Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY · ORGANIZED IN 1935

VOLUME 6
NUMBER 1
FEBRUARY
1941

PSYCHOMETRIKA, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued six times a year, on February 15, April 15, June 15, August 15, October 15, and December 15.

FEBRUARY, 1941, VOLUME 6, NUMBER 1

Office of Publication: 2539 Briarcliffe Avenue, Cincinnati, Ohio. Entered as second class matter September 17, 1940, at the post office at Colorado Springs, Colorado, under the act of March 3, 1879.

Subscription Price:

To non-members, the subscription price is $5.00 per volume of four issues. Members of the Psychometric Society pay annual dues of $5.00, of which $4.50 is in payment of a subscription to *Psychometrika.* Student members of the Psychometric Society pay annual dues of $3.00, of which $2.70 is in payment for the journal.

The subscription price to libraries and other institutions is $10.00 per year; this price includes one extra copy of each issue.

Applications for membership and student membership in the Psychometric Society should be sent to

JOHN C. FLANAGAN
Chairman of the Membership Committee
15 Amsterdam Ave.,
New York, N.Y.

Membership dues and subscriptions are payable to

IRVING J. LORGE
Institute of Educational Research
Columbia University
New York, New York

(Continued on the back inside cover page)

Psychometrika

VOLUME VI—1941

JANUARY-DECEMBER

PUBLISHED BI-MONTHLY

By THE PSYCHOMETRIC SOCIETY

AT 23 WEST COLORADO AVENUE

COLORADO SPRINGS, COLORADO

S

T

T

O

G

N

M

A

MU

VO

Psychometrika

CONTENTS

VOLUME SIX FEBRUARY 1941 NUMBER ONE

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

SCORING FORMULAS AND PROBABILITY CONSIDERATIONS

ALEXANDER CALANDRA
DEPARTMENT OF CHEMISTRY, BROOKLYN COLLEGE

Bayes' theorem of inverse probability is made the basis of a general equation for scoring objective examinations. The equation so obtained is evaluated by assuming a binomial distribution of examinee knowledge and guessing tendency. A graphical illustration of the application of this equation to a hypothetical test situation is presented. The limitations inherent in the use of Bayes' theorem make it inadvisable to recommend the practical use of the equation unless future experimental evidence indicates an increase in scoring validity which more than compensates for the increase in scoring difficulty.

During the past two decades a number of statistical studies have been undertaken for the purpose of determining the relative merits of the formulas which have been proposed for the scoring of objective examinations. As a result of these studies, it has become evident that the formulas which seemed to be based on the soundest theoretical considerations were hardly, if at all, superior to those of empirical origin. Now, while this apparent failure of the theory of probability to provide a reasonably superior approach to the development of scoring formulas has brought forth a number of explanations whose validity has always been considered an open question, there has been little or no attempt to re-examine the theoretical foundations of the formulas. This situation has prompted the writer to present the following development of a general method of scoring based upon considerations which may be more appropriate than those suggested up to this point.

The purpose in developing any scoring formula is the derivation of that statistical function which will give the best estimate of the number of correct responses which were not the result of guessing. The problem which this raises is of the general type dealing with events which are the result of the operation of known and unknown causes, and can be classified as "a priori" or "a posteriori" depending on whether it involves the calculation of the most probable nature of the event from a knowledge of the known causes, or the calculation of the most probable extent of the known causes from a knowledge of the event.

Problems in "a priori" probability are amenable to simple solutions of unquestioned validity. Thus on the basis of the most elemen-

tary considerations the answer to the problem of determining the most probable raw score of a student who knows 50 items on a true-false test and guesses at an additional 30, is given by the expression

$$50 + 30/2 = 65 .$$

Problems in "a posteriori" probability are often of considerable complexity and even when solvable sometimes yield ambiguous or even meaningless results. Thus the problem of determining the most probable value of the number of items known by a student who obtains a raw score of 25 correct items and 50 incorrect items yields a meaningless solution when it is assumed that the number of responses guessed correctly is equal to the number of responses guessed incorrectly. Most scoring problems involve questions of "a posteriori" probability and the formulas which have been derived for their solution are based on assumptions analogous to the one indicated above. It is possible however, to avoid such assumptions by an application of Bayes' theorem of inverse probabilities* and to obtain a general solution which is applicable to a variety of objective questions.

According to Bayes' theorem the probability $P_R{}^k$ that an event R was the result of a cause k is given by the expression,

$$P_R{}^k = \frac{P_k P_k^R}{\Sigma P_z P_z{}^R}, \tag{1}$$

where

$P_k =$ the probability that the cause k exists.
$P_z =$ the probability that any cause z exists.
$P_k{}^R =$ the probability that the operation of cause k will give rise to the event R.
$P_z{}^R =$ the probability that the operation of any cause z will give rise to the event R.

The proof of (1) can be found in most textbooks dealing with probability.†

By extending Bayes' theorem it is possible to develop a general scoring formula in terms of the following definitions:

$n =$ the total number of items on the examination.
$R =$ the number of items the examinee answers correctly.
$W =$ the number of items the examinee answers incorrectly.

* The use of Bayes' theorem as a basis for scoring formulas was suggested by the writer in his unpublished doctorate dissertation at the School of Education of New York University.

† Fisher, A. The mathematical theory of probabilities. New York: The MacMillan Co., 1930.

$y =$ the number of items the examinee attempts;

$$y = R + W$$

$x =$ the number of items known to the examinee.

$P_x^n =$ the probability that an examinee who knows x items exists in the group taking the test.

$P_{y-x}^{n-x} =$ the probability that an examinee who knows x items will guess at $y-x$ additional items.

$P_{R-x}^{y-x} =$ the probability that an examinee who guesses at $y-x$ items will obtain $(R-x)$ correct guess responses, i.e., a total of R correct responses.

$P_{R,y}^{\ x} =$ the probability that a raw score of R correct items out of a total of y responses originated from a knowledge of x items.

Bayes' theorem can be used to calculate $P_{R,y}^{\ x}$ by means of the substitutions

$$P_{R,y}^{\ x} = P_R^{\ k} \ ;$$

$$P_x^n = P_z = P_k \ ;$$

$$P_{y-x}^{n-x} \cdot P_{r-x}^{y-x} = P_z^{\ R} = P_k^{\ R} \ .$$

These substitutions yield

$$P_{R,y}^{\ x} = \frac{P_x^n \cdot P_{y-x}^{n-x} \cdot P_{R-x}^{y-x}}{\sum_{x=0}^{R} P_x^n \cdot P_{y-x}^{n-x} \cdot P_{R-x}^{y-x}} \ . \tag{2}$$

By the use of (2) and the theorem of the mean value, the score S, which can be assigned to an examinee who attempts y items and answers R correctly is given by the equation

$$S = \sum_{x=0}^{R} x \, P_{R,y}^{\ x} \ . \tag{3}$$

According to equation (3) the best score which can be assigned to a paper containing R correct responses out of a total of y responses is the mean value of x for all such papers.

Combining (2) and (3),

$$S = \frac{\sum_{x=0}^{R} x \cdot P_x^n \cdot P_{y-x}^{n-x} \cdot P_{R-x}^{y-x}}{\sum_{x=0}^{R} P_x^n \cdot P_{y-x}^{n-x} \cdot P_{R-x}^{y-x}} \ . \tag{4}$$

The probabilities P_x^n , P_{y-x}^{n-x} P_{R-x}^{y-x} can be evaluated by considering them to arise from binomial distributions generated by the corresponding probabilities p_1 , p_2 , and the probability p_3 , which are defined as follows,

p_1 = the probability that an examinee knows any given item.

p_2 = the probability that an examinee will guess at any given item.

p_3 = the probability that an examinee who guesses at an item will guess it correctly.

These definitions make possible the following evaluations:

$$P_x^n = C_x^n \, (p_1)^x \, (1-p_1)^{n-x} = \frac{n!}{(n-x)!\,(x)!} (p_1)^x \, (1-p_1)^{n-x} ;$$

$$P_{y-x}^{n-x} = \frac{(n-x)!\,(p_2)^{y-x}\,(1-p_2)^{n-y}}{(n-x-y+x)!\,(y-x)!} ;$$

$$P_{R-x}^{y-x} = \frac{(y-x)!\,(p_3)^{R-x}\,(1-p_3)^{y-R}}{(y-R)!\,(R-x)!} .$$

These expressions can be simplified by the use of the following substitutions:

$$k_1 = \frac{p_1}{1-p_1} , \quad k_2 = \frac{p_2}{1-p_2} , \quad \text{and} \quad k_3 = \frac{p_3}{1-p_3} ,$$

yielding the relationships

$$P_x^n = \frac{n!\,(1-p_1)^n\,(k_1)^x}{(n-x)!\,(x)!} ;$$

$$P_{y-x}^{n-x} = \frac{(n-x)!\,(1-p_2)^n\,(k_2)^y\,(p_2)^{-x}}{(n-y)!\,(y-x)!} ;$$

$$P_{R-x}^{y-x} = \frac{(y-x)!\,(1-p_3)^y\,(k_3)^R\,(p_3)^{-x}}{(y-R)!\,(R-x)!} .$$

As a result of the foregoing relationships,

$$\Sigma P_x^n \cdot P_{y-x}^{n-x} \cdot P_{R-x}^{y-x}$$

$$= \Sigma \frac{n!(n-x)!(y-x)!(1-p_1)^n(1-p_2)^n(1-p_3)^y(k_1)^x(k_2)^y(k_3)^R(p_2)^{-x}(p_3)^{-x}}{(n-x)!(y-x)!(n-y)!(y-R)!(R-x)!(x)!} . \tag{5}$$

In determining the value of S for any single examination paper, it should be noted that n , R , y , p_1 , p_2 , p_3 , (k_1), (k_2), (k_3) are all con-

stants and that only those functions containing x are variable. It is therefore possible to make the substitution in (5)

$$\frac{n!(1-p_1)^n(1-p_2)^n(1-p_3)^y(k_2)^y(k_3)^R}{(n-y)!\,(y-R)!}=K$$

and obtain as a result

$$\sum_{x=0}^{R} P_x{}^n \cdot P_{\substack{n-x\\ y-x}} \cdot P_{\substack{y-x\\ R-x}} = K\sum_{x=0}^{R}\frac{(k_1)^x(p_2)^{-x}(p_3)^{-x}}{(R-x)!(x)!}\,.$$

But

$$(k_1)^x\,(p_2)^{-x}\,(p_3)^{-x}=\left[\frac{k_1}{p_2\cdot p_3}\right]^x.$$

For the purpose of algebraic convenience, let

$$\frac{k_1}{p_2\,p_3}=Z\,.$$

Then

$$\sum_{x=0}^{R} P_x{}^n \cdot P_{\substack{n-x\\ y-x}} \cdot P_{\substack{y-x\\ R-x}} = K\sum_{x=0}^{R}\frac{(Z)^x}{(R-x)!(x)!}\equiv KQ_D\,. \tag{6}$$

Expansion of

$$Q_D=\sum_{x=0}^{R}\frac{Z^x}{(R-x)!(x)!}$$

$$=\frac{1}{R!\,0!}+\frac{Z}{(R-1)!1!}+\frac{Z^2}{(R-2)!2!}+\frac{Z^3}{(R-3)!3!}+\cdots.$$

Multiplying by $R!$,

$$Q_DR!=1+\frac{ZR}{1!}+\frac{Z^2R(R-1)}{2!}+\frac{Z^3(R)(R-1)(R-2)}{3!}+\cdots.$$

Comparison of this series with the expansion of $(a+b)^n$ shows that

$$Q_DR!=(1+Z)^R\,; \tag{7}$$

$$Q_D=\sum_{x=0}^{R}\frac{Z^x}{(R-x)!(x)!}=\frac{(1+Z)^R}{R!}\,. \tag{8}$$

Substituting (8) in (6),

$$\sum_{x=0}^{R} P_x{}^n \cdot P_{\substack{n-x\\ y-x}} \cdot P_{\substack{y-x\\ R-x}} = \frac{K(1+Z)^R}{R!}\,. \tag{9}$$

The numerator of (4) can be evaluated by an analogous process, all steps being identical, leading to the expression

$$\sum_{x=0}^{R} x\, P_x{}^n\, P_{y-x}^{n-x}\, P_{R-x}^{y-x} = K \sum_{x=0}^{R} \frac{x(Z)^x}{(R-x)\,!\,(x)\,!} \equiv KQ_N \,. \tag{10}$$

By expansion,

$$\sum_{x=0}^{R} \frac{x(Z)^x}{(R-x)\,!\,(x)\,!} = \frac{Z}{(R-1)\,!\,(1)\,!} + \frac{2Z^2}{(R-2)\,!\,(2)\,!} + \frac{3Z^3}{(R-3)\,!\,(3)\,!}$$

or (11)

$$Q_N = \frac{Z}{(R-1)\,!\,(1)\,!} + \frac{Z^2}{(R-2)\,!\,(1)\,!} + \frac{Z^3}{(R-3)\,!\,(2)\,!} + \frac{Z^4}{(R-4)\,!\,(3)\,!}$$

and

$$\frac{Q_N(R-1)\,!}{Z} = 1 + \frac{Z(R-1)}{(1)\,!} + \frac{Z^2(R-1)(R-2)}{2!} + \frac{Z^3(R-1)(R-2)(R-3)}{3!} + \cdots .$$

But by comparison of the resulting series with the expansion of $(a+b)^n$,

$$\frac{Q_N(R-1)\,!}{Z} = (1+Z)^{R-1}$$

and

$$Q_N = \frac{Z(1+Z)^{R-1}}{(R-1)\,!} = \frac{Z(1+Z)^{R-1}R}{R!} \,. \tag{12}$$

Combining (4), (9), and (12) yields

$$S = \frac{K\,Z(1+Z)^{R-1}R}{K(1+Z)^R} = \frac{ZR}{1+Z} = \frac{R}{1+\dfrac{1}{Z}} \,;$$

$$S = \frac{R}{1+\left[\dfrac{1}{p_1}-1\right]p_2p_3} \,. \tag{13}$$

In applying (13) to the scoring of any set of objective items, it is necessary to evaluate p_1, p_2, p_3. For a set of true-false questions this can be done as follows:

Evaluation of p_1:

In deriving (13) it was assumed that a binomial distribution of examinee knowledge existed. The value of p for this distribution can be estimated from the raw scores made on the examination by the group taking the test and is equal to

$$p_1 = \frac{\Sigma R - \Sigma W}{\Sigma n},$$

where

ΣR is the sum of all the correct responses by all the examinees,
ΣW is the sum of all the incorrect responses by all the examinees,
Σn is the sum of all the examination items on all the papers.

Evaluation of p_2:

The derivation of (13) assumed that a binomial distribution of guessing tendency existed. The value of p for this distribution can be obtained from the raw scores of all students who answered exactly $R+W=K_y$ items, and is equal to

$$p_2 = \frac{\text{the number of items guessed at}}{\text{the number of items available for guessing}}$$

or

$$p_2 = \frac{\text{the number of items attempted} - \text{the number of items known}}{\text{the number of items on the test} - \text{the number of items known}}.$$

This can be estimated as

$$p_2 = \frac{(\Sigma R) + (\Sigma W) - [(\Sigma R) - (\Sigma W)]}{(\Sigma n) - [(\Sigma R) - (\Sigma W)]} = \frac{2(\Sigma W)}{(\Sigma n) - (\Sigma R) + (\Sigma W)},$$

where

(ΣR) is the sum of all the correct responses made by all the examinees answering exactly K_y items, and corresponding definitions hold for the other terms.

Evaluation of p_3:

From the nature of a set of true-false questions,

$$p_3 = 1/2.$$

From these evaluations, and equation (13)

$$S = \frac{R}{1 + \left[\frac{\Sigma n - \Sigma R + \Sigma W}{\Sigma R - \Sigma W}\right]\left[\frac{(\Sigma W)}{(\Sigma n) - (\Sigma R) + (\Sigma W)}\right]}. \qquad (14)$$

The significance of equation (14) can be illustrated by its application to the following hypothetical test situation. A true-false test of 20 items is submitted to a large group of examinees. It is desired to assign a score to an examinee who answers only 16 items and makes 12 correct responses.

Curve I This represents the assumed binomial frequency distribution of examinee knowledge. In this curve $n=20$, and p_1, which should be calculated from the examination data, has been taken as $1/2$.

Curve II This represents the frequency distribution of knowledge among those examinees who answer exactly K_y items. In this curve K_y is 16, and p_2, which should be calculated from the data, has been taken as $1/2$.

Curve III This represents the frequency distribution of knowledge among those examinees who answer exactly 16 items and obtain exactly 12 correct responses. The abscissa corresponding to the ordinate which bisects the area under curve II is given by S. This can be calculated from the graph or from the equation and is in both cases equal to 9.8.

Equations (4), (13), and (14) are presented chiefly as a matter of theoretical interest. The practical use of these equations is not recommended unless adequate experimental evidence indicates an increase in scoring validity which more than compensates for the increase in scoring difficulty. The writer hopes to make available in a later paper a discussion of the practical and theoretical limitations of these equations.

The writer takes pleasure in acknowledging his indebtedness to Edward E. Cureton, John C. Flanagan, Charles T. Molloy, and Paul V. West for their very helpful discussions of the subject matter in this article.

An illustration of an application of Bayes' Theorem to a scoring problem.

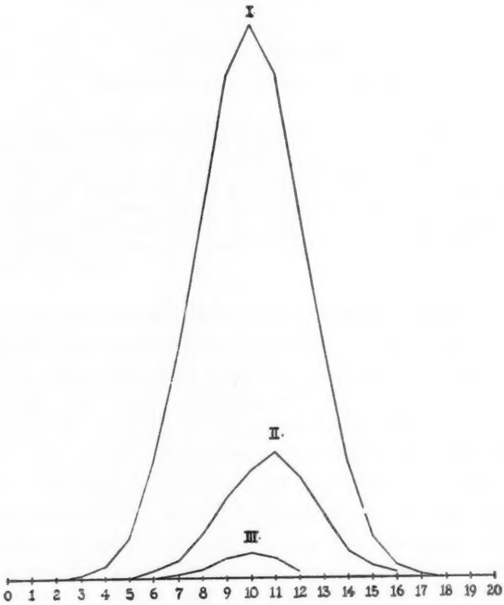

FIGURE 1

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

THE PHI COEFFICIENT AND CHI SQUARE AS INDICES OF ITEM VALIDITY

J. P. GUILFORD
UNIVERSITY OF SOUTHERN CALIFORNIA

Two new methods of item analysis are described. One involves the computation of the ϕ coefficient (correlation of a fourfold point distribution) and the other involves chi square. The only data required are the proportions of passing individuals in the upper and lower criterion groups, for the determination of ϕ, and in addition, N, for the determination of chi square. Abacs are presented for graphic solution of the two indices of validity, and tests of significance are provided.

Many of the devices for determining item validity are based upon the principle of the correlation between an item and some criterion variable. It is also common practice to employ dichotomous classifications in both the item variable and the criterion variable, using the proportions of passing individuals in the high and low criterion groups. The two criterion groups, furthermore, are equal populations—upper and lower quarters, 27 per cents, or halves of the entire criterion population. The two methods to be described are applicable when this situation obtains. The one method, which requires the computation of a ϕ coefficient, is based upon the correlation principle. The other method, which involves the computation of chi square, rests upon the principle of divergence of the proportions of passing individuals among the high and low criterion groups from a purely random or chance distribution. Both indices of validity have the logical advantage that they do not require the assumption of a continuous distribution in either the item or criterion variable although both are applicable when one or both distributions are continuous.* There are times when the criterion variable in particular represents a division into two discrete classes, as for example the two sexes when items for masculinity-femininity are being validated.

The Phi Test of Validity

The customary formula for the computation of ϕ is given as

$$\phi = \frac{\alpha\delta - \beta\gamma}{\sqrt{pqp'q'}} \tag{1}$$

* I am indebted to Mr. H. M. Cox for a critical review of the ideas set forth in this paper.

in which α, β, γ and δ are the proportions occupying the four cells of a fourfold table,

p is the sum of the first row and equals $\alpha + \beta$,
q is the sum of the second row, $\gamma + \delta$,
p' is the sum of the first column, or $\alpha + \gamma$,
and q' is the sum of the second column, or $\beta + \delta$.

When applied to the correlation of an item with a criterion, as under the conditions previously specified, the computation of ϕ is much simplified for the reason that $p' = q' = .5$, since the populations in upper and lower criterion groups are equal in size. The fourfold table appears as follows, if we make certain substitutions in symbols:

	U	L	
P	p'_u α	p'_l β	p
F	q'_u γ	q'_l δ	q
	.5	.5	1.00

In this table U and L stand for the upper and lower criterion groups. P and F stand for the passing and failing populations. The symbol p'_u stands for the proportion of the entire criterion population (including both upper and lower groups) who are both in the upper group and pass the item. The other primed symbols are to be interpreted in analogous manner.

With these symbols the formula for ϕ becomes

$$\phi = \frac{p'_u q'_l - p'_l q'_u}{.5\sqrt{pq}} . \tag{2}$$

From the fourfold table given above,

$$q'_u = .5 - p'_u$$

and

$$q'_l = .5 - p'_l .$$

Substituting these for q'_u and q'_l in (2),

$$\phi = \frac{.5p'_u - p'_u p'_l - .5p'_l + p'_u p'_l}{.5\sqrt{pq}}$$

$$= \frac{.5(p'_u - p'_l)}{.5\sqrt{pq}} \tag{3}$$

$$= \frac{p'_u - p'_l}{\sqrt{pq}}.$$

But it is usually more convenient to derive from the test data the proportions of the upper and lower groups alone who pass an item. Let these proportions be designated by p_u and p_l, respectively. Since $p_u = 2p'_u$ and $p_l = 2p'_l$,

$$p'_u - p'_l = \frac{p_u - p_l}{2},$$

and therefore, substituting in (3),

$$\phi = \frac{p_u - p_l}{2\sqrt{pq}}, \tag{4}$$

which is the formula proposed for practical use. It requires only the knowledge of the proportion of successes in the upper and lower criterion groups (p_u and p_l) and the proportions of successes and failures in the two groups combined (p and q). The constant p is derivable directly from p_u and p_l, being given by

$$p = \frac{p_u + p_l}{2},$$

and $q = l - p$.

It may be pointed out in passing that the simple method of item analysis which employs merely the index $(p_u - p_l)$ ingores entirely the dispersion of individuals on the item variable. The expression $\sqrt{pq}$ in the denominator of (4) gives the standard deviation of the item provided we assume a symmetrical dispersion of the item. The S. D. of the item is at a maximum when $p = q = .5$ and approaches zero when p approaches 0 or 1.00. The ϕ index of validity of an item is seen to be directly proportional to the difference in proportion of successes in the two criterion groups, but at the same time it is inversely proportional to the variability of the item. If ϕ is adopted as the index of validity, then it follows that the same numerical difference $(p_u - p_l)$ is more significant when p approaches 0 or 1.00 and least significant when $p = q = .50$.

In order to facilitate the computation of ϕ, an abac was constructed as shown in Fig. 1.* In using the graphic solution of ϕ,

* I am indebted to N.Y.A. assistance in the preparation of the diagrams in Figures 1 and 2.

only the values of p_u and p_l need be known. p_u is to be sought on the ordinate and p_l on the abscissa. The intersection of the two lines corresponding to them gives the value of ϕ which is sought.

The method of constructing the abac may be of interest. Beginning with the equation (4) and multiplying through by $\sqrt{pq}$, we have

$$\phi\sqrt{pq} = \frac{p_u - p_l}{2} = (p_u - p) = (p - p_l) . \tag{5}$$

By letting ϕ take on constant values in turn, one can readily determine the lines of equal ϕ. For example, if we are interested in the line for $\phi = .10$, letting p take on various values we can find a set of coordinates which will determine the line. When $p = .9$, $\sqrt{pq} = .3$, and $\phi\sqrt{pq} = .03$. From equation (5), p_u then equals .93 and p_l equals .87. Other coordinates are similarly computed.

The lower right-hand part of the abac is a mirror reflection of the upper-left part, with p_u and p_l having interchanged roles. The diagram is given in complete form with all negative ϕ's not only because an occasional item correlates negatively with the criterion but also because the abac has general application to problems other than that of item analysis, remembering, of course, that it applies only when $p' = q' = .50$, and that $p_u = 2p'_u$ and $p_l = 2p'_l$. Presumably other abacs could be constructed on the same principle for varying values of p' and q', and a set of graphic charts having a role similar to that of the Thurstone tetrachoric diagrams (1) would be the result. It is doubtful, however, whether such charts would have sufficient general application to justify their publication.

There is a question as to the comparability of a ϕ coefficient under different conditions of separation of upper and lower criterion groups. Exclusion of the middle segment of an entire criterion population, when the variable is continuous, usually has the effect of augmenting the numerical index of correlation. But so long as a constant rule of division is followed, upper versus lower quarters, for example, the ϕ coefficients should be comparable. Phi coefficients from situations having varied degrees of exclusion of middle cases could perhaps be made numerically consistent, but the writer has not attempted to provide this refinement. As a matter of fact, owing to varying ranges of talent or of trait difference in different populations, even the consistent use of the same division, even of halves, may not guarantee numerical consistency for any kind of index of validity.

The question of size of significant ϕ coefficients for varying val-

ues of N will be discussed following the presentation of the second method, for it depends upon the size of chi square.

The Chi-Square Method

With all the variety of procedures for determining item validity it is surprising that mention of the chi-square technique has not more frequently appeared. The direct relation of chi square to ϕ immediately suggests the use of this statistic which is so prominent in the newer statistics of small samples. The fact that tests of significance are so readily forthcoming when chi square is computed makes this statistic appealing in connection with item analysis. The relation to ϕ is given by the equation

$$\chi^2 = N\phi^2 = \frac{N(p_u - p_l)^2}{4pq}, \tag{6}$$

to use the particular formula for ϕ which applies to the special case of item analysis (formula (4)).

It would be possible, of course, to compute values of chi square for all items, either using the formula just given, or by means of nomographs designed for the purpose. The parameter N is an additional element in the situation, however, and it necessitates a series of abacs rather than just one as in the case of the computation of ϕ. Instead of this extended computational practice, the writer recommends a simple test of significance of item validity for varying values of N. The size of N (the number of individuals in the upper and lower criterion groups combined) is usually chosen for convenience as some multiple of 10. An N of 50 is probably the minimal practical value with which to make a test of validity of items and an N of 400 is probably the maximal practical value. The abac (see Fig. 2) was prepared to show the lines of "significant" and "very significant" chi squares for N's of 50, 100, 200, and 400.* The locus of the coordinates for a constant value of N was determined as follows. Multiplying equation (6) through by $4pq/N$, we have

$$\frac{4pq\chi^2}{N} = (p_u - p_l)^2. \tag{7}$$

When $N = 50$,

$$.08pq\chi^2 = (p_u - p_l)^2. \tag{8}$$

* The designations "significant" and "very significant" have Fisher's usual meaning. A "significant" chi square could occur in a truly random situation 5 times in 100 ($P = .05$) and a "very significant" chi square could occur similarly only once in a hundred times ($P = .01$).

Reference to Fisher's table of chi squares (2) shows that for one degree of freedom, which obtains in a fourfold table, it requires a chi square of 6.635 to be called very significant. Substituting this value in equation (8), we have

$$.5308pq = (p_u - p_l)^2 .$$

Taking the positive square roots we have,

$$.7285\sqrt{pq} = p_u - p_l . \tag{9}$$

Letting p take on different values, sets of coordinates for a constant chi square of 6.635 and an N of 50 can be computed. For example, if $p = .8$, $\sqrt{pq} = .4$. Then from (9), $p_u - p_l = .2914$. But $p_u - p_l = 2(p_u - p) = 2(p - p_l)$. Therefore, the two coordinates we seek are:

$$p_u = .800 + \frac{.2914}{2} = .946$$

and

$$p_l = .800 - \frac{.2914}{2} = .654 ,$$

to use three significant figures.

Other pairs of coordinates were computed in a similar manner, for other values of p and for N's of 100, 200, and 400. In using the abac in Fig. 2, any item whose coordinates p_u and p_l locate it above and to the left of the curved line appropriate to the N of our criterion population may be regarded as significantly valid or as very significantly valid according as its chi square is greater than 3.841 ($P = .05$) or 6.635 ($P = .01$), respectively. The line of zero chi square would of course be defined by the equation $p_u = p_l$ and it would form the diagonal of the chart. It was found empirically that the lines for a chi square of 3.841 were practically identical with those for a chi square of 6.635 when N is doubled. From this it follows that in practice, doubling N when p_u and p_l remain constant has the effect of lowering the probability of a random situation from .05 to .01, or of moving an item from the category of "significantly valid" to that of "very significantly valid."

The use of the chi-square chart in Fig 2 has the limitation of merely discriminating between significantly versus non-significantly valid items when it is used as already described. It would not place items in the order of their validity. Rank order of the valid items may be approximated, however, by plotting each item as a point within the diagram and noting the relative diagonal distance from the line of least significance.

Because of the close relationship between chi square and ϕ, tests of significance of ϕ can be readily inferred from the tests of significance of chi square. In other words, we can determine the least "significant" and least "very significant" values of ϕ for different values of N. It is simply a matter of determining the values of ϕ that are equivalent to chi squares of 3.841 and 6.635, respectively, for different values of N. A brief list of these significant ϕ's is given in Table 1. The table has a small number of values of N, but their range is

TABLE 1

Minimum Values of ϕ That are Significant and Very Significant for Various Sizes of N

	$X^2 = 3.841, P = .05$	$X^2 = 6.635, P = .01$
N	ϕ	ϕ
20	.438	.576
30	.358	.470
40	.310	.407
50	.277	.364
70	.234	.308
100	.196	.258
150	.160	.210
200	.139	.182
300	.112	.150
400	.098	.129
600	.080	.105

more extensive than is needed in ordinary item-analysis problems with the idea that it may be useful elsewhere.* Its use in connection with ϕ coefficients should proceed with caution, however, particularly when dealing with populations from which middle cases have been excluded. In continuous distributions it is probably only strictly applicable when upper and lower halves are included. When the table is applicable, it naturally obviates the necessity of computing standard errors of ϕ.

REFERENCES

1. Chesire, L., Saffir, M., and Thurstone, L. L. Computing diagrams for the tetrachoric correlation coefficient. Chicago: Univ. Chicago Bookstore, 1933.
2. Fisher, R. A. Statistical methods for research workers. Edinburgh: Oliver & Boyd, 1938.

* Even when upper and lower quarters of a criterion population do not exactly equal multiples of 10, a random sampling will make them do so, for general convenience as well as for convenience in using the abacs and the table presented here.

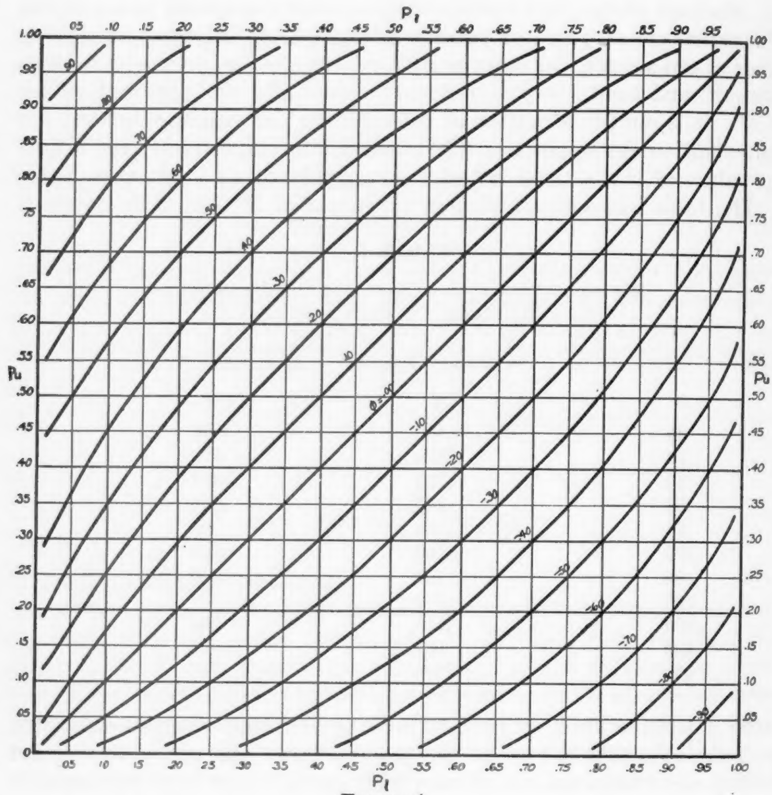

FIGURE 1

Abac for the graphic estimation of ϕ as an index of item validity. p_u = the proportion of the upper criterion group who pass the item, and p_l is the similar proportion of the lower criterion group.

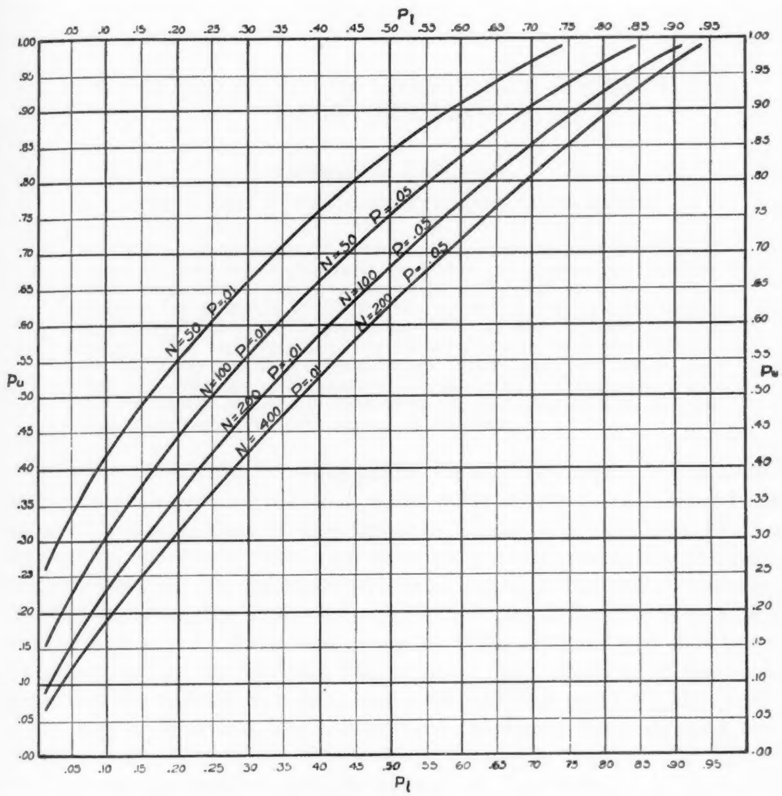

Figure 2

Abac for the graphic determination of items with significant and very significant validity. p_u = the proportion of the upper criterion group who pass the item, and p_l is the similar proportion of the lower criterion group.

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

THE APPLICATION OF SHEPPARD'S CORRECTION FOR GROUPING

GEORGE A. FERGUSON
DEPARTMENT OF EDUCATION, UNIVERSITY OF EDINBURGH

This paper attempts to show in a non-mathematical way the influence of grouping on standard deviations and correlations, and advances empirical evidence to illustrate with what accuracy values corrected for grouping by Sheppard's correction approximate to values obtained from ungrouped data when the distributions are continuous. This enquiry gained its initial stimulus from the observation that many standard deviations and correlations reported by students of psychology and education are uncorrected for grouping and that frequently errors attributable to the grouping of data are not small when compared with errors of sampling.

In the calculation of statistical measures from grouped data the values of each variate within a given class-interval are assigned the value of the mid-point of that interval. Thus in the calculation of a correlation coefficient from such data we are not calculating the relationship between the continuous variates x and y, but rather the relationship between the midpoints of certain class-intervals into which the variates x and y have been grouped. With distributions that taper off to zero at the extremities, the point of concentration of the variate is not the mid-point of the class-interval but a point slightly nearer the mean. Thus statistical measures calculated from the odd moments remain almost uninfluenced by grouping, because the errors made by the assumption that the scores are concentrated at the midpoint of each interval tend to balance on both sides of the mean, while with the even moments the errors do not balance but add together.

Grouping error increases the size of the uncorrected standard deviations, and reduces the size of the uncorrected correlations. The usual formula for correcting a standard deviation for grouping is as follows:

$$s = \sqrt{\bar{s}^2 - \frac{i^2}{12}}\,; \tag{1}$$

where s, $\bar{s}$ are the corrected and uncorrected estimates, respectively, of the standard deviation, and i is the class-interval.

The correction to be applied to a correlation coefficient for grouping depends on the observation that when the distributions of the two

correlated variates tail off at the extremities the quantity $\bar{r}_{xy}\ \bar{s}_x\ \bar{s}_y$ is independent of the class-interval used. It immediately follows from this observation that

$$r_{xy} = \frac{\bar{r}_{xy}\ \bar{s}_x\ \bar{s}_y}{s_x\ s_y}, \tag{2}$$

where $\bar{r}_{xy}$ and r_{xy} are the uncorrected and corrected values of the correlation between x and y. Since, however, $\bar{r}_{xy}\ \bar{s}_x\ \bar{s}_y = \sum xy/N$, the usual product-moment formula for a correlation coefficient corrected for grouping may be written as follows:

$$r_{xy} = \frac{\sum xy}{N\sqrt{(\bar{s}_x^2 - \frac{i_x^2}{12})(\bar{s}_y^2 - \frac{i_y^2}{12})}}, \tag{3}$$

where i_x and i_y represent the class-intervals of x and y, respectively. When correlation coefficients are calculated by the diagonal adding method, the formula for a corrected coefficient becomes

$$r_{xy} = \frac{H + V - D}{2\sqrt{(H - \frac{N}{12})(V - \frac{N}{12})}}, \tag{4}$$

where H, V, and D represent the sum of the squares of the deviations from the mean values of x, y, and $x-y$, respectively.

R. A. Fisher* has pointed out that in averaging correlation coefficients the values of z should be obtained from uncorrected values of r and that a correction equivalent to the average correction of the averaged values of r should be added to the resulting coefficient.

The corrected value of r is always larger than the uncorrected value of r. The larger the value of r the larger the absolute value of the correction to be made for grouping. The relative value of the correction is constant, given constant values for the standard deviations of the variates correlated. The size of the correction is independent of N, the number of cases. Errors introduced by using uncorrected values of r when r is large are much more significant than errors resulting from a corresponding grouping when r is small. Not only is the absolute discrepancy between the uncorrected and the corrected value of r greater when r is large, but small differences between large correlations represent a much greater difference in the degree of relationship between the variates correlated than equivalent differences between small coefficients.

* R. A. Fisher. Statistical methods for research workers, p. 211. London: Oliver and Boyd, 1938.

EXPERIMENTAL. To estimate the accuracy with which values corrected for grouping approximate to values obtained from ungrouped data when the distribution is regarded as continuous, the I.Q.'s of 952 children on two intelligence tests were plotted on a grid with a class-interval of unity. The two distributions of scores were approximately normal. The standard deviations of the two variables, and the correlation between them were calculated. The class-interval was then successively increased by telescoping, at it were, the original grid, and further standard deviations and correlations were calculated with class-intervals of 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, and 20.

Table 1 gives the uncorrected and corrected standard deviations for variable x at different units of class interval, and the number of arrays upon which each measure is based. The corrected standard deviation with a class-interval of unity is taken as the standard, and the deviations from this standard of the uncorrected and corrected standard deviations, calculated at each step interval, are given in columns d_1 and d_2, respectively. It will be observed that the uncorrected standard deviation with a class-interval of unity is the same as would have been obtained from ungrouped data. This value is, however, corrected on the basis of the assumption that the distribution is theoretically continuous.

TABLE 1

class-interval	no. of arrays	$\bar{s}_x$ (uncorrected)	s_x (corrected)	d_1	d_2
1	60	12.1550	12.1516	.0034	.0000
2	30	12.1549	12.1412	.0033	—.0104
3	20	12.1634	12.1325	.0118	—.0191
4	15	12.1740	12.1191	.0224	—.0325
5	12	12.1175	12.0313	.0341	—.1203
6	10	12.1836	12.0599	.0320	—.0917
7	9	12.3123	12.1452	.1607	—.0064
8	8	12.4592	12.2433	.3076	.0917
9	7	12.6432	12.3734	.4916	.2218
10	6	12.4806	12.1421	.3290	.0095
12	5	12.4620	11.9708	.3104	—.1808
14	5	12.6512	11.9883	.4996	—.1633
16	4	12.8747	12.0177	.7231	—.1339
18	4	13.1897	12.1230	1.0381	—.0286
20	3	13.3611	12.0493	1.2095	—.1023

Table 2 furnishes corresponding data for variable y. These data indicate that the application of Sheppard's correction results in an estimate of the standard deviation closely approximating to the value

that would have obtained from an ungrouped continuous variate. Certain relatively large discrepancies in the corrected values occasionally appear and are due to the purely arbitrary nature of the points fixed as the top of the last class interval and the bottom of the first.

TABLE 2

class-interval	no. of arrays	$\bar{s}_x$ (uncorrected)	s_x (corrected)	d_1	d_2
1	55	11.2309	11.2272	.0037	.0000
2	28	11.2563	11.2416	.0291	.0144
3	19	11.2518	11.2184	.0246	—.0088
4	14	11.3123	11.2523	.0851	.0260
5	11	11.3570	11.2645	.1298	.0373
6	10	11.3988	11.2664	.1716	.0392
7	8	11.3421	11.1595	.1149	.0677
8	7	11.4128	11.1768	.1856	.0504
9	7	11.5848	11.2896	.3576	.0624
10	6	11.5273	11.1600	.3001	—.0872
12	5	11.8006	11.2807	.5634	.0535
14	4	11.5885	10.8608	.3613	—.3664
16	4	11.7920	10.8498	.5648	—.3774
18	4	12.5132	11.3834	1.2860	.1562
20	3	12.5510	11.1442	1.3238	—.0830

In the calculation of correlation coefficients by the diagonal adding method, adding along one diagonal of the correlation grid yields the distribution of the sum of the variates, while adding along the other diagonal yields the distribution of the difference between the variates. Thus if we wish to calculate the standard deviation of variation in I.Q. between test and retest, instead of finding the actual difference in I.Q. for every child and making a distribution of these differences, we may find the standard deviation of difference in I.Q. directly from the appropriate diagonal distribution. A peculiarity exists, however, in the grouping of the diagonal distribution which makes the standard deviation of $x - y$ calculated from it greater than the standard deviation of $x - y$ calculated from the distribution obtained by subtracting the appropriate values of y from x and grouping the differences with class-interval equal to that of x and y in the original grid. We may readily correct for this peculiarity since the variance of the diagonal distribution is greater than the variance of a distribution of actual differences of the same class-interval by an amount equal to $\frac{i^2}{12}$. Thus if the latter standard deviation is corrected for grouping once, the former must be corrected twice. This

point is capable of further illustration by reference to the formula

$$\tilde{s}_{(x-y)} = \tilde{s}_x^2 + \tilde{s}_y^2 - 2\,\tilde{r}_{xy}\tilde{s}_x\tilde{s}_y\,.$$

Since the term $\tilde{r}_{xy}\,\tilde{s}_x\,\tilde{s}_y$ is invariant, it is apparent that the term $\tilde{s}^2_{(x-y)}$ must be corrected twice if $\tilde{s}_x^2$ and $\tilde{s}_y^2$ are each corrected and the equation is to be satisfied. Furthermore, this simple observation indicates why there is no correction in any portion of the numerator of formula (4). The numerator of this formula really reads

$$(H - \frac{N}{12}) + (V - \frac{N}{12}) - (D - \frac{2N}{12})\ ,$$

so that the corrections cancel one another, leaving the numerator invariant.

To illustrate the foregoing discussion, the standard deviation of the diagonal distribution was calculated at different class-intervals, and these values uncorrected, corrected once, and corrected twice are given in Table 3. The standard deviation of the difference with class-interval unity is taken as the standard value, and the deviations d_1, d_2, d_3 of the standard deviations at different class-intervals, uncorrected, corrected once, and corrected twice, from this standard value are given. It is apparent from an examination of the data in this table that twice Sheppard's correction is the correction required.

TABLE 3

class-interval	no. of arrays	$\tilde{s}_{(x-y)}$ (uncorrected)	$s_{(x-y)}$ (corrected once)	$s_{(x-y)}$ (corrected twice)	d_1	d_2	d_3
1	35	5.9401	5.9331	5.9261	.0140	.0070	.0000
2	18	5.9662	5.9382	5.9101	.0401	.0121	.0160
3	12	6.0219	5.9592	5.8960	.0958	.0331	—.0301
4	10	6.1348	6.0251	5.9134	.2087	.0990	—.0157
5	10	6.2517	6.0828	5.9091	.3256	.1567	—.0170
6	7	6.2382	5.9929	5.7371	.3121	.0668	—.1890
7	7	6.5170	6.1958	5.8570	.5909	.2697	—.0691
8	6	6.6952	6.2843	5.8428	.7691	.3582	—.0833
9	5	7.0182	6.5196	5.9794	1.0921	.5935	.0533
10	6	7.2845	6.6836	6.0330	1.3584	.7575	.1069
12	5	7.4767	6.6258	5.6481	1.5506	.6997	—.2780
14	5	8.3318	7.2860	6.0624	2.4057	1.3599	.1363
16	3	8.5042	7.1406	5.4456	2.5781	1.2145	—.4805
18	3	9.1499	7.5313	5.4516	3.2238	1.6052	—.4745
20	3	9.9576	8.1130	5.6997	4.0315	2.1869	—.2264

The correlations between the variates x and y were also calculated at different units of class-interval. These values are given in Table 4. Here again the corrected value with class-interval unity is

taken as the standard, and the deviations d_1 and d_2 of the obtained and corrected values of r from this standard are calculated. A very substantial decrease in the value of r with decrease in the number of arrays into which the variates are grouped is observed. The discrepancy between the uncorrected and corrected values of r is such as to furnish sound support to the conclusion that correlation coefficients must be corrected for grouping if accurate statistics are desired. These data are indicative that Sheppard's correction furnishes a remarkably accurate estimate of the correlation that would have obtained from ungrouped data with continuous variates.

TABLE 4

class-interval	no. of arrays x	no. of arrays y	$\bar{r}_{xy}$ (uncorrected)	r_{xy} (corrected)	d_1	d_2
1	60	55	.8739	.8744	.0005	.0000
2	30	28	.8729	.8750	.0015	.0006
3	20	19	.8706	.8754	.0038	.0010
4	15	14	.8661	.8746	.0083	.0002
5	12	11	.8601	.8733	.0143	—.0011
6	10	10	.8621	.8812	.0123	.0068
7	9	8	.8513	.8771	.0231	.0027
8	8	7	.8462	.8793	.0282	.0049
9	7	7	.8357	.8762	.0387	.0018
10	6	6	.8187	.8692	.0557	.0052
12	5	5	.8114	.8836	.0630	.0092
14	5	4	.7671	.8638	.1073	—.0106
16	4	4	.7656	.8914	.1088	.0170
18	4	4	.7478	.8943	.1266	.0199
20	3	3	.7063	.8821	.1681	.0077

In order to examine the functioning of Sheppard's correction with a small value of r, a new grid was drawn up with 1828 cases. Values of r were found as before at successive class-intervals. Table 5 gives values of r uncorrected and corrected for different class-intervals. The deviations of the uncorrected and corrected values, respectively, from a standard value .3672 are given in columns d_1 and d_2. The number of arrays is given, in this case the number of arrays of the x variable being equal to the number of arrays of the y variable for each value of r.

It will be observed that the d_1 column of Table 4 is in every case greater than the d_1 column of Table 5, illustrating that the larger the value of r the larger the absolute value of Sheppard's correction, and emphasizing that correcting for grouping is of much more importance when r is large than when r is small. Examination of the d_2 columns of Tables 4 and 5 shows that Sheppard's correction furnishes

TABLE 5

class-interval	no. of arrays	$\bar{r}_{xy}$ (uncorrected)	r_{xy} (corrected)	d_1	d_2
1	60	.3670	.3672	.0002	.0000
class-	no. of	r_{xy}	r_{xy}		
2	30	.3663	.3672	.0009	.0000
3	20	.3648	.3668	.0024	.0004
4	15	.3632	.3667	.0040	.0005
5	12	.3613	.3667	.0059	.0005
6	10	.3581	.3660	.0091	.0012
7	9	.3548	.3653	.0124	.0019
8	8	.3520	.3658	.0152	.0014
9	7	.3514	.3685	.0158	—.0013
10	6	.3457	.3669	.0215	.0003
12	5	.3340	.3634	.0332	.0038
14	5	.3452	.3873	.0220	—.0101
16	4	.3134	.3616	.0538	.0056
18	4	.3112	.3758	.0560	—.0086
20	3	.2729	.3423	.0943	.0249

a remarkably accurate estimate of the correlation that would have obtained from ungrouped data with continuous variates. Furthermore, if there is reason to believe that the distributions of the two correlated variables have "high contact," some work can be avoided by using a coarse grouping with a small number of arrays and correcting for grouping. Tables 4 and 5 show that quite accurate results can be obtained with as few as six arrays. With fewer than six arrays the purely arbitrary position of the class-intervals will in most cases lead to slight discrepancies in the corrected value of r.

SUMMARY. If the distributions of variates used in statistical work taper off to zero at the extremities, the use of Sheppard's correction furnishes accurate estimates of the standard deviations and correlations that would have resulted from the use of ungrouped data. Correcting a correlation coefficient for grouping is essential when the grouping is coarse and the number of arrays is large. Otherwise inaccurate statistics will result. The discrepancies found in small correlations due to failure to correct for grouping are of less importance. Sheppard's correction for grouping should be used whenever the grouping error cannot be considered small in relation to errors of random sampling.

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

ON THE RECTILINEAR PREDICTION OF OBLIQUE FACTORS

HARRY H. HARMAN
UNIVERSITY OF CHICAGO

The general problem of estimating correlated or uncorrelated factors is treated. It is specifically indicated wherein the prediction of oblique factors differs from that of orthogonal factors. A shortened method of estimation of correlated factors is developed.

Factor analysis is concerned primarily with two basic problems: (1) the linear resolution of a set of variables in terms of hypothetical factors; and (2) the description of these factors in terms of the observed variables. There are many and sundry methods for the solution of the first of these problems.* A common feature of most of these methods, however, is that for a set of n variables the final form of solution reveals $m(<n)$ common factors and n unique factors. Since the total number of factors exceeds the number of variables, the value of any particular factor for a given individual can only be estimated from the observed values of the variables. The best prediction, in the least-square sense, is that obtained by the ordinary regression method. It is thus apparent that the solution of the second problem is in the nature of a linear regression equation.

Let it be assumed that a set of n variables has been analyzed in terms of m common factors and n unique factors, as follows:

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + a_jU_j \quad (j = 1, 2, \cdots, n). \tag{1}$$

The observed variables and the factors are standardized over the sample of N individuals, so that the standard value of the variable z_j for an individual i may be written explicitly in the form:

$$z_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} + \cdots + a_{jm}F_{mi} + a_{ji}U_{ji} \begin{pmatrix} j = 1, 2, \cdots, n \\ i = 1, 2, \cdots, N \end{pmatrix}. \tag{1'}$$

In general, the form of equation (1) is preferred, the secondary subscript for the individual values being understood. A set of equations of the type (1) is said to be a *factor pattern*. In this definition no assumption is made as to the correlations among the factors. It is usually postulated that the unique factors are uncorrelated among themselves and with all common factors. The common factors, how-

* For a description of the methods leading to the "preferred types" of factor solutions see (1, Part II).

ever, may or may not be correlated. A factor pattern may be written conveniently in the form of a matrix equation, namely,

$$Z = MF\,, \tag{2}$$

where the column vectors

$$Z = \{z_1\, z_2 \cdots z_n\}$$

$$F = \{F_1 \cdots F_m\, U_1 \cdots U_n\}$$

represent the variables and factors, respectively, and the matrix M, which consists of the matrix A of common factor coefficients and the matrix U of unique factor coefficients, is defined by

$$M = \|A\ U\| = \begin{Vmatrix} a_{11}\, a_{12} \cdots a_{1m} & a_1 & 0 & \cdots & 0 \\ a_{21}\, a_{22} \cdots a_{2m} & 0 & a_2 & \cdots & 0 \\ \cdot\ \cdot\ \cdot\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1}\, a_{n2} \cdots a_{nm} & 0 & 0 & \cdots & a_n \end{Vmatrix}.$$

When there can be no confusion, the *pattern matrix* M of factor coefficients may be referred to as the factor pattern.

The rectilinear prediction of any Factor F_s involves the determination of the coefficients in the regression equation

$$\bar{F}_s = \beta_{s1} z_1 + \cdots + \beta_{sj} z_j + \cdots + \beta_{sn} z_n \,(s = 1\,, 2\,, \cdots, m)\,, \tag{3}$$

where the $(n-1)$ secondary subscripts have been omitted from the β's for convenience. A similar equation can be written for any one of the unique factors. It will be convenient to have the symbol F_s represent any one of the complete set of factors $F_1\,, F_2\,, \cdots, F_m\,, U_1\,, U_2\,, \cdots, U_n$. In general, then,

$$\bar{F} = BZ\,, \tag{4}$$

where $\bar{F}$ represents the matrix of factor estimates and B the matrix of regression coefficients.

By the ordinary method for obtaining the regression of one variable on n others, any factor F_s can be estimated when the correlations of the variables with this factor and among themselves are known. Setting

$$D = \begin{Vmatrix} 1 & r_{F_s z_1} & r_{F_s z_2} & \cdots & r_{F_s z_n} \\ r_{z_1 F_s} & 1 & r_{12} & \cdots & r_{1n} \\ r_{z_2 F_s} & r_{21} & 1 & \cdots & r_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{z_n F_s} & r_{n1} & r_{n2} & \cdots & 1 \end{Vmatrix},$$

the regression coefficient of z_j in the estimation of F_s is given by

$$\beta_{sj} = -\frac{D_{sj}}{D_{ss}} , \tag{5}$$

where D_{ss} is the minor of the first element and D_{sj} is the cofactor of $r_{F_s z_j}$ in D. It may be noted that D is merely the matrix R of the observed correlations bordered by the correlations of the variables with F_s. Then $D_{ss} = R$ and the determinants D_{sj} can be expressed in terms of the cofactors of the original correlation matrix. Thus the values (5) may be written explicitly in the form:

$$\beta_{sj} = \frac{1}{R} \left[r_{z_1 F_s} R_{1j} + r_{z_2 F_s} R_{2j} + \cdots + r_{z_n F_s} R_{nj} \right] , \tag{6}$$

where R_{kj} is the cofactor of r_{kj} in the correlation determinant R.

The regression equation (3) for F_s may then be written:

$$\bar{F}_s = \| t_{1s}\, t_{2s} \cdots t_{ns} \| R^{-1} Z , \tag{7}$$

where $t_{js} = r_{z_j F_s}$. In particular, if F_s is one of the unique factors, say U_j, this equation becomes:

$$\bar{U}_j = \| 0\ 0 \cdots t_j \cdots 0 \| R^{-1} Z , \tag{8}$$

or, in expanded form,

$$\bar{U}_j = \frac{t_j}{R} \left(R_{1j} z_1 + R_{2j} z_2 + \cdots + R_{nj} z_n \right) , \tag{8'}$$

where $t_j = r_{z_j U_j}$. The matrix equation (4) for the prediction of the entire set of factors finally becomes:

$$\bar{F} = T' R^{-1} Z , \tag{9}$$

where T' is the transpose of the *factor structure*,

$$T = \left\| \begin{matrix} t_{11} & t_{12} & \cdots & t_{1m} & t_1 & 0 & \cdots & 0 \\ t_{21} & t_{22} & \cdots & t_{2m} & 0 & t_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ t_{n1} & t_{n2} & \cdots & t_{nm} & 0 & 0 & \cdots & t_n \end{matrix} \right\| .$$

All the factors, whether they be correlated or not, can be estimated by means of equation (9). When the factors are mutually orthogonal, however, it can be shown (2, p. 324) that the elements of M are equal to the corresponding elements of T, i.e., the pattern and structure coincide. Then equation (9) may be written as follows:

$$\bar{F} = M'R^{-1}Z \qquad \text{(uncorrelated factors)}. \tag{10}$$

On the other hand, if the factors *are* correlated equation (10) does not apply, and the distinction between pattern and structure must be clearly understood (1, Section 2.4 and Chapter XI).

The use of equation (10) has been the conventional method for estimating orthogonal factors. Spearman's explicit formula (3, Appendix) for a single general factor is a special case of this equation, and the methods employed by Thurstone (4) and Thomson (5) for the case of several common factors are adaptations of this formula. Thomson also gave the explicit formula (8′) for the estimation of unique factors.

It will be noted that in the application of formula (10) the evaluation of R^{-1} is involved. This may be very laborious if the number of variables is large. For this reason, the author (6) has developed several approximations to this formula, which involve the grouping of certain variables to reduce the order of the matrix whose inverse is required. More recently, Ledermann (7) and Guttman (8) have developed a shortened method for the estimation of orthogonal factors which involves the inverse of an $m \times m$ matrix instead of the inverse of an $n \times n$ matrix and which is just as accurate as the longer method. The effect of this method is to replace the observed correlations by those reproduced (or computed) from the factor pattern.

Before the shortened method is generalized to the case of oblique factors, a formula will be derived for the prediction of correlated factors which explicitly employs the pattern matrix M. Let the matrix of correlations among the common factors be defined by:

$$\phi = \left\| \begin{matrix} 1 & r_{F_1F_2} & \cdots & r_{F_1F_m} \\ r_{F_2F_1} & 1 & \cdots & r_{F_2F_m} \\ \cdot & \cdot & \cdot & \cdot \\ r_{F_mF_1} & r_{F_mF_2} & \cdots & 1 \end{matrix} \right\|,$$

so that the matrix of correlations among all the factors is

$$\Phi = \left\| \begin{matrix} \phi & 0 \\ 0 & I \end{matrix} \right\|.$$

By a fundamental theorem of factor analysis (1, 2, 9), the matrix of correlations of the variables can be expressed in the form:

$$R = M\Phi M'. \tag{11}$$

Then, noting that

$$M \Phi M' R^{-1} = RR^{-1} = I \, , \tag{12}$$

and premultiplying the members of (2) by the members of (12) there results:

$$M \Phi M' R^{-1} Z = MF \, .$$

The premultiplier M, although not square, may be dropped from both sides, and, again using a bar to denote estimated values, the equation finally becomes:

$$\bar{F} = \Phi M' R^{-1} Z \, . \tag{13}$$

This equation reduces to (10) when all the factors are uncorrelated.

Since the unique factors are of minor interest in factor analysis, it is convenient to write equation (13) for the common factors only, namely,

$$\bar{f} = \phi A' R^{-1} Z \, , \tag{14}$$

where the small f is used to denote the column vector $\{F_1 \; F_2 \cdots F_m\}$. In case the common factors are uncorrelated, this equation reduces to:

$$\bar{f} = A' R^{-1} Z \qquad \text{(uncorrelated factors)}. \tag{15}$$

In formulas (13) and (14) the pattern values are employed instead of the structure values.

Now the method of Ledermann and Guttman will be generalized to the case of oblique common factors. From (11),

$$R = \| A \; U \| \cdot \left\| \begin{matrix} \phi & 0 \\ 0 & I \end{matrix} \right\| \cdot \left\| \begin{matrix} A' \\ U \end{matrix} \right\| = \| A \phi \; U \| \cdot \left\| \begin{matrix} A' \\ U \end{matrix} \right\| = A \phi A' + U^2 \, . \tag{16}$$

Then, premultiplying R by $A'U^{-2}$ the following expression is obtained, which ultimately leads to the desired result:

$$A'U^{-2}R = A'U^{-2}(A\phi A' + U^2) = (A'U^{-2}A\phi + I)A'$$

$$A'U^{-2}R = (I + K)A' \, , \tag{17}$$

where $K = A'U^{-2}A\phi$. Premultiplying both members of (17) by $(I + K)^{-1}$ and postmultiplying by R^{-1}, this equation becomes

$$(I + K)^{-1} A' U^{-2} = A' R^{-1} \, . \tag{18}$$

Then, substituting (18) into (14), there finally results:

$$\bar{f} = \phi (I + K)^{-1} A' U^{-2} Z \, . \tag{19}$$

This equation indicates that, even when the factors are correlated, only the inverse of an $m \times m$ matrix is required.

To show the general character of formula (19), it is convenient to write it in another form. Premultiply both sides of (19) by $[\phi(I + K)^{-1}]^{-1}$, we get:

$$(I + K)\phi^{-1}\bar{f} = A'U^{-2}Z . \tag{20}$$

By noting the order of each constituent matrix in (20), it will be observed that each of the two matrices resulting from the multiplications contains m rows and one column. This matrix equation thus represents a system of m algebraic equations, obtained by setting the corresponding elements equal to each other. The elements of the matrices in the right-hand member of (20) are obtained simply enough, but the expression on the left appears to be rather complex. This can be simplified, however. Substituting the definition of K in the premultiplier of $\bar{f}$, this expression becomes:

$$(I + K)\phi^{-1} = (I + A'U^{-2}A\phi)\phi^{-1} = (\phi^{-1} + A'U^{-2}A). \tag{21}$$

The system of m equations for the estimation of the common factors may finally be written in the form:

$$(\phi^{-1} + A'U^{-2}A)\bar{f} = A'U^{-2}Z . \tag{22}$$

From this equation it is evident that, first, the inverse of the $m \times m$ matrix of intercorrelations of factors is calculated, to this is added the $m \times m$ matrix $(A'U^{-2}A)$, and then the inverse of this composite $m \times m$ matrix is required in order to obtain the explicit equations for the estimation of oblique factors. The expression (22) includes Ledermann's formula (7, eq. 11) as a special case, namely, when ϕ is the identity matrix. There is also some similarity between this formula and the one derived by Bartlett (10, eq. 4). Of course, since Bartlett has assumed a different principle of estimation than that employed in this paper, his formula is not a special instance of (22). The analogy exists in the fact that if the term ϕ^{-1} is dropped in equation (22), it becomes identical with Bartlett's formula.

REFERENCES

1. Holzinger, Karl J. and Harman, Harry H. Factor analysis: A synthesis of factorial methods (in Press).
2. Holzinger, Karl J. and Harman, Harry H. Relationships between factors obtained from certain analyses. *J. educ. Psychol.*, 1937, 28, 321-345.
3. Spearman, Charles. The abilities of man. New York: Macmillan, 1927.
4. Thurstone, L. L. The vectors of mind. Chicago: Univ. Chicago Press, 1935.
5. Thomson, Godfrey H. Some points of mathematical technique in the fac-

torial analysis of ability. *J. educ. Psychol.*, 1936, **27**, 37-54.

6. Harman, Harry H. Systems of regression equations for the estimation of factors. *J. educ. Psychol.*, 1938, **29**, 431-441.
7. Ledermann, Walter. On a shortened method of estimation of mental factors by regression. *Psychometrika*, 1939, **4**, 109-116.
8. Guttman, Louis. Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 1940, **5**, 75-99.
9. Tucker, Ledyard R. The role of correlated factors in factor analysis. *Psychometrika*, 1940, **5**, 141-152.
10. Bartlett, M. S. The statistical conception of mental factors. *Brit. J. Psychol.*, 1937, **28**, 97-104.

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

GENETIC EMERGENCE OF FACTOR SPECIFICITY

T. W. RICHARDS
SAMUEL S. FELS RESEARCH INSTITUTE
YELLOW SPRINGS, OHIO

Mental test data of Chrysostomm and of Garrett, Bryan, and Perl are reinvestigated in order to determine the shifts in test clusters with chronological age. It is found that the factors under survey tend to become more independent with increasing age.

One theory of the genesis of mental organization is that mental abilities become more specific as children grow older. According to this notion, it might be supposed that fairly discrete or uncorrelated factors at a given age level might be expected to overlap at an earlier age level. Two studies of test intercorrelations at succeeding age levels were used to determine if possible the progressive shifts in test clusters with chronological age. Garrett, Bryan, and Perl (2) used ten tests on nine-, twelve- and fifteen-year-old-children, both boys and girls. Six matrices of intercorrelations were thus obtained. The tests used are described in some detail in the report (2). At least five were "memory" tests, and four were "non-memory." Chrysostom (1) used nine tests with Belgian school boys in the fourth, fifth, and sixth grades. His material has not, as far as I know, been published except as an abstract (1). Chrysostom's tests consisted of French adaptations of five *Stanford Achievement Tests* (*paragraph meaning, sentence meaning, word meaning, arithmetic problems, arithmetic computation*), a *substitution* test taken from Woodworth and Wells' *Association Tests* and from the *Army Mental Tests, following directions* from Burt's *Mental and Scholastic Tests,* and an *information* test, taken from Haggerty's *Intelligence Examination Delta,* for grades 3 to 9. These were described as above in personal communication (Dec., 1939).

The numbers of cases used in the various groups were as follows:

	Garrett, et al		*Chrysostom (boys)*	
	Boys	*Girls*		
nine years	108	117	4th grade	56
twelve years	96	100	5th grade	57
fifteen years	102	123	6th grade	54

Rotation of axes for these nine matrices involved extraction of

factors for Chrysostom's material, and recalculating much of the Garrett material. In the case of each matrix, two factors seemed significant. However, for Chrysostom's fifth-grade group and Garrett's fifteen-year-old boys it was necessary to use the third extracted factor rotated with the first, rather than k_1 and k_2 as in every other case.

It appeared to me that for the oldest age groups in each case, Chrysostom's sixth-grade and Garrett's two fifteen-year-old groups, two factors were established by rather clear clusters. Although naming these factors seemed secondary to tracing the formation of clusters genetically, I might say that Chrysostom's factors seemed to be a "verbal" factor, heaviest in tests 1, 2, 3, and 8 (see Table 1), and a "directions," or ability to follow directions or solve problems, found in tests 4, 5, 6, and 7. The factors in Garrett's material seem to me to be a memory factor found in tests 6, 7, 8, 9, and 10, and a (non-memory) factor, perhaps speed, found in tests 1, 2, 3, 4, and 5.

Factor loadings appear in Tables 1 and 2.

Figures 1, 2, and 3 illustrate the shift in clusters with chronolog-

TABLE 1

Unrotated and Rotated Values from Chrysostom (1)

Unrotated Loadings

		Grade						
		4		5			6	
	Test	K_1	K_2	K_1	K_2	K_3	K_1	K_2
1	Reading Paragraph	.785	—.226	.638	—.103	—.462	.675	.293
2	Reading Sentences	.697	—.226	.561	—.120	—.149	.598	.353
3	Word Meaning	.735	—.414	.732	—.255	—.118	.671	.508
4	Arithmetic Problems	.695	.258	.871	.267	.102	.722	—.356
5	Arithmetic Computation	.659	.186	.718	.357	.143	.612	—.504
6	Substitution	.519	.340	.431	—.135	.389	.390	—.301
7	Directions	.555	.428	.684	—.291	.311	.576	—.265
8	Opposites	.578	—.333	.586	.285	—.242	.673	.326

Rotated Loadings

		Grade									
		4			5				6		
	Test	*I*	*II*	h^2	*I*	*II*	K_2	h^2	*I*	*II*	h^2
1	Reading Paragraph	.74	.35	.67	.78	.13	—.10	.63	.70	.24	.54
2	Reading Sentences	.67	.29	.54	.50	.29	—.12	.35	.68	.14	.48
3	Word Meaning	.83	.17	.71	.60	.44	—.26	.62	.84	.08	.71
4	Arithmetic Problems	.35	.65	.55	.54	.69	.27	.84	.29	.75	.65
5	Arithmetic Computation	.37	.57	.47	.40	.61	.36	.66	.11	.79	.63
6	Substitution	.16	.59	.39	.02	.58	—.14	.36	.08	.49	.25
7	Directions	.13	.68	.49	.26	.70	—.29	.65	.25	.58	.40
8	Opposites	.65	.13	.45	.58	.25	.29	.48	.73	.20	.56

TABLE 2

Rotated Values from Garrett, Bryan, and Perl(2)

Nine Years

		Boys			*Girls*		
		I	*II*	h^2	*I*	*II*	h^2
1	Making gates	.39	.17	.18	.31	.45	.29
2	Vocabulary	.30	.53	.37	.49	.30	.33
3	Arithmetic	.73	.30	.62	.62	.31	.47
4	Paper Form Board	.36	—.03	.12	.63	.09	.40
5	Logical Prose	.37	.62	.52	.40	.65	.57
6	Word-Word	—.04	.62	.39	.21	.51	.30
7	Word Retention	.37	.56	.44	.53	.21	.31
8	Digit Span	.34	.15	.14	.37	.49	.37
10	Geometric Forms	.42	.19	.21	.07	.47	.22
11	Objects	.56	.61	.67	.15	.59	.36

Twelve Years

		Boys			*Girls*		
		I	*II*	h^2	*I*	*II*	h^2
1	Making Gates	.40	.26	.23	.32	.20	.14
2	Vocabulary	.68	.25	.55	.82	—.06	.67
3	Arithmetic	.67	.43	.64	.59	.32	.44
4	Paper Form Board	.39	.57	.46	.57	.22	.37
5	Logical Prose	.76	.29	.66	.61	.14	.40
6	Word-Word	.13	.46	.23	.35	.20	.16
7	Word Retention	.06	.57	.32	—.09	.60	.36
8	Digit Span	.48	.04	.22	.06	.49	.25
10	Geometric Forms	.17	.35	.15	.16	.10	.04
11	Objects	.12	.69	.47	.23	.50	.30

Fifteen Years

		Boys				*Girls*		
		I	*II*	h^2	K_2	*I*	*II*	h^2
1	Making gates	.47	.10	.29	—.26	.47	.08	.23
2	Vocabulary	.52	.37	.54	.38	.70	.07	.49
3	Arithmetic	.46	.16	.31	.27	.72	.00	.52
4	Paper Form Board	.39	—.05	.18	—.14	.46	.04	.22
5	Logical Prose	.23	.16	.11	.19	.54	.10	.30
6	Word-Word	.20	.18	.12	.22	.16	.55	.32
7	Word Retention	.05	.42	.35	—.42	.00	.57	.32
8	Digit Span	.17	.45	.26	—.20	.16	.19	.06
10	Geometric Forms	—.04	.17	.06	.18	.18	.24	.09
11	Objects	.11	.28	.15	.24	.06	.49	.24

ical age. In the case of the Chrysostom material, it is seen that the sixth-grade clusters (white) are most discrete, with the fourth-grade clusters (dotted shading) less discrete but more so than the fifth-grade material. Possibly the fifth-grade material represented a sampling of some sort less comparable with the other two grades than

these were with each other. The charts for the Garrett et al material show that for the fifteen-year-old girls there is a shift in discreteness from the ninth through the twelfth to the fifteenth year. The case for the boys is less clear, perhaps, but similar. Here there is consid-

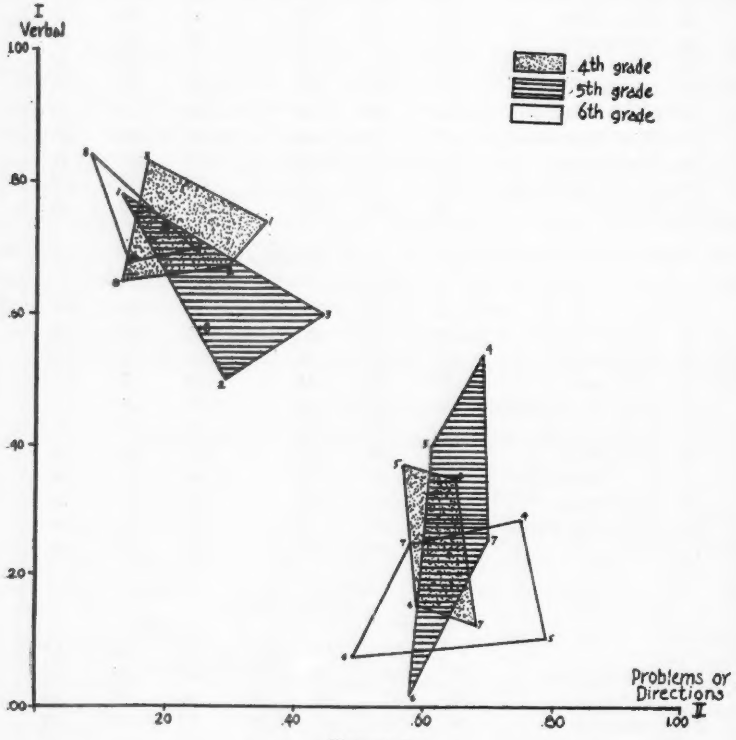

FIGURE 1

Factors I & II from Chrysostom material, for fourth-, fifth-, and sixth-grade boys.

erable overlapping at the nine-year level between the clusters; at twelve years there is no overlapping but the memory cluster gets over into the non-memory area definitely, because of test 8. At fifteen years there is no overlapping, but total variance is low.

The two most confusing levels, — Garrett's fifteen-year-old boys where total variance was low, and Chrysostom's fifth grade where there was a reversion in the age tendency from purity to impurity— were those cases in which it was necessary to extract a third factor to get the homologue to the second factor in each other case.

Since, as far as I know, these studies are the only ones showing a tendency for factors to become more independent or orthogonal with increasing age, it seemed that a brief report of them might be of interest.

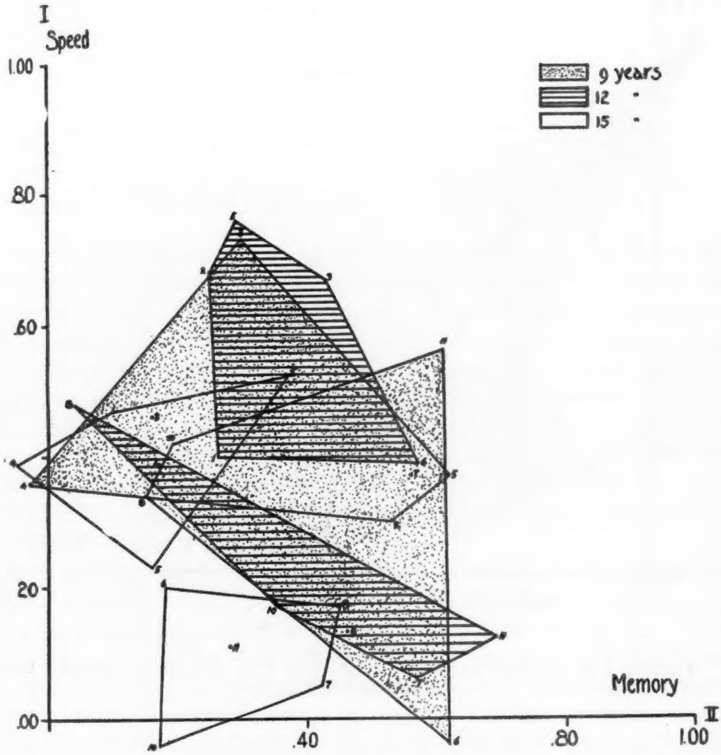

FIGURE 2
Factors I & II from Garrett, Bryan, and Perl, for nine-, twelve-, and fifteen-year-old girls.

REFERENCES

1. Chrysostom, F. L'organization des traits mentaux avec l'age. *Annales de l'Assoc. Canadienne-Francoise pour l'avancement des sciences,* 1938, 4, 124.
2. Garrett, Henry E., Bryan, Alive I. & Perl, Ruth E. The age factor in mental organization. *Archives of Psychology,* No. 176, 1935, pp. 31.

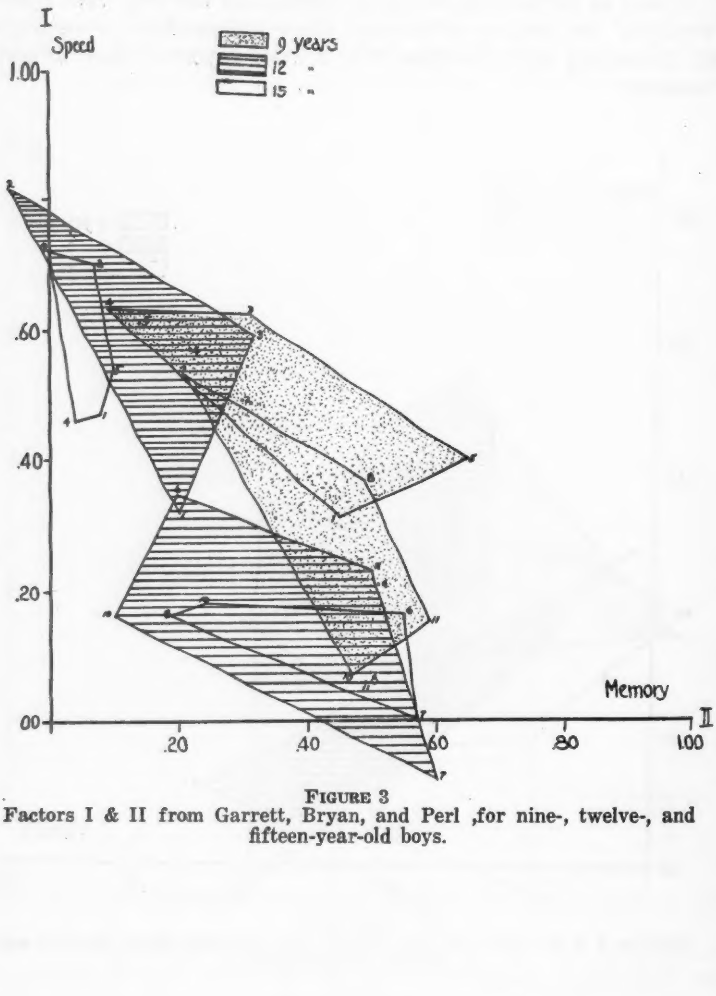

FIGURE 3
Factors I & II from Garrett, Bryan, and Perl ,for nine-, twelve-, and fifteen-year-old boys.

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

NOTE ON THE MATHEMATICAL THEORY OF INTERACTION OF SOCIAL CLASSES

N. RASHEVSKY
THE UNIVERSITY OF CHICAGO

In continuation of previous studies, the interreaction of two social classes is investigated from a somewhat different point of view. An exchange of results of activities of the two classes is considered, and a mathematical approach is outlined based on the use of such psychological concepts as the satisfaction function.

In previous papers, (2, 3, 4) we have studied mathematically different types of interaction of social classes. In one of the papers (2) we discussed interactions based mainly on psychological principles, while in others (2, 3) we considered an interaction that may perhaps best be called "economic," though it also involved sociological and psychological mechanisms. In the present paper we shall discuss a possible mathematical approach from a still different point, which, while psychological in its essence, may be applied as well to interactions that resemble more the "economic" type. This approach consists in a generalization to social classes of some concepts that have been studied for the case of single individuals.

Consider again two coexisting classes I and II. Let both classes, amongst other activities, perform two given activities A and B, producing correspondingly per unit time some results a and b. Those results may be considered as "commodities," either of a material nature, like food, or of a more abstract nature, like knowledge of some sort, that may be communicated to others. Let the amounts of the results of activities produced by the first class be a_1 and b_1, by the second a_2 and b_2. It may happen that a_1 is rather large while a_2 is small, and at the same time b_1 small while b_2 is large. In that case an exchange of "commodities" will take place, the first class receiving some b from the second and giving in return some a.

To determine the character of that exchange, we shall use, as has been done by other authors (1, 5) the concept of satisfaction, which we may apply to a class as a whole, if it consists of approximately similar individuals. That satisfaction, as a psychological quantity, may be actually, though indirectly, measured and discussed quantitatively has been shown by L. L. Thurstone (5). Thurstone comes

to the conclusion, derived from psychophysical experimental evidence, that the satisfaction varies logarithmically with the amount of commodity possessed, and he assumes moreover that for several commodities the satisfactions are simply additive. However, he makes some reservations as to the generality of the logarithmic relation. For the present general discussion, we shall not make any special assumptions about the shape of the satisfaction function. Moreover, we shall consider the satisfaction not in terms of the quantities of commodities possessed, but in terms of the rates of productions of those commodities. This is psychologically legitimate, for one may derive a greater satisfaction from producing or receiving per unit time more commodity. We shall speak therefore of "production" of a commodity by a class, if that commodity is received from outside, for instance from the other class.

We assume that for any individual of class I there is a satisfaction function $s_1(x , y)$, where x and y are the amounts of the two commodities in question received per unit time. Similarly for class II we have $s_2(x , y)$. If N_1 and N_2 are the numbers of individuals in class I and class II, respectively, we shall define $S_1(x , y) = N_1 s_1(x , y)$ and $S_2(x , y) = N_2 s_2(x , y)$ as the satisfaction function for classes I and II, respectively. We shall put

$$\frac{\partial S_1}{\partial x} = X_1(x , y) \ ; \quad \frac{\partial S_1}{\partial y} = Y_1(x , y) \ ;$$
$$\frac{\partial S_2}{\partial x} = X_2(x , y) \ ; \quad \frac{\partial S_2}{\partial y} = Y_2(x , y) \ . \tag{1}$$

If each individual in a class, and therefore the class as a whole, agrees to such an exchange, for which his satisfaction has the largest possible value, and if this exchange goes on in such a way that for a unit of x always the same number of units of y are given, then we may calculate the results of such an exchange by using formulae developed by G. Evans (1, pp. 125–128). Denoting by x_1 and y_1 the rates of production of x and y in class II when the exchange is operating, and by x_2 and y_2 corresponding quantities for class I, we have for the determination of x_1, y_1, x_2, and y_2 the following equations

$$\frac{X_1(x_1 , y_1)}{X_2(a_1 + a_2 - x_1 , b_1 + b_2 - y_1)} = \frac{Y_1(x_1 , y_1)}{Y_2(a_1 + a_2 - x_1 , b_1 + b_2 - y_1)} \, . \tag{2}$$

$$x_1 + x_2 = a_1 + a_2 \; ;$$
$$y_1 + y_2 = b_1 + b_2 \, . \tag{3}$$

Equations (2) and (3) express x_1, y_1, x_2, and y_2 in terms of a_1, b_1, a_2, and b_2

$$x_1 = u_1(a_1, b_1, a_2, b_2) \; ; \quad x_2 = u_2(a_1, b_1, a_2, b_2) \; ;$$
$$y_1 = v_1(a_1, b_1, a_2, b_2) \; ; \quad y_2 = v_2(a_1, b_1, a_2, b_2) \, . \tag{4}$$

Depending on the choice of the functions S_1 and S_2, a different distribution of rates of production will be obtained. It is possible that while $a_1 + b_1 > a_2 + b_2$, yet $x_1 + y_1 < x_2 + y_2$.

The quantities a_1, b_1, a_2, and b_2 refer to the class as a whole. The corresponding quantities per individual shall be denoted by a'_1, a'_2, b'_1, and b'_2. Thus $a_1 = N_1 a'_1$, etc.

Consider the theoretical case in which at the time $t = 0$ class I consists of n_{01} individuals capable of activity A only, while class II is composed of n_{02} individuals capable of activity B only. We shall refer to them as individuals of type A and type B. An individual of type A is characterized by

$$a' > 0 \, , b' = 0 \, . \tag{5}$$

An individual of type B is characterized by

$$a' = 0 \, , b' > 0 \, . \tag{6}$$

We thus have at $t = 0$

$$a'_1 > 0 \; ; \quad b'_1 = 0 \; ; \quad a'_2 = 0 \; ; \quad b'_2 > 0 \, . \tag{7}$$

Now consider a variation of the composition of each class due to a mechanism discussed in a previous paper (2), and based on a dissimilarity of offspring and parents. Denoting by $n^A{}_I$ the number of individuals in the class I who are performing activity A, by $n^B{}_I$ the number of individuals in class I performing activity B, and by $n^A{}_{II}$ and $n^B{}_{II}$ the corresponding quantities for class II, we find for the variations of these quantities with respect to time equations of the form of equations (43) and (46) of a previous paper (2). We thus have

$$n^A{}_I = n^A{}_I(t) \; ; \quad n^B{}_I = n^B{}_I(t) \; ; \quad n^A{}_{II} = n^A{}_{II}(t) \; ; \quad n^B{}_{II} = n^B{}_{II}(t) \, . \tag{8}$$

When class I is composed of individuals performing both activities A and B, then the value of a_1 will be less than for the case of a class consisting of individuals performing only activity A. On the

other hand, b_1 will now not be zero. We have now

$$a_1 = n^A{}_I a'_1 \;; \quad b_1 = n^B{}_I b'_2 \,. \tag{9}$$

Similarly, for the second class

$$a_2 = n^A{}_{II}\, a'_2 \;; \quad b_2 = n^B{}_{II}\, b'_2 \,. \tag{10}$$

Since the n's are functions of time, a_1, b_1, a_2, and b_2 will also be known functions of time. Individuals of the two different types will have different satisfaction functions, $s_A(x, y)$ and $s_B(x, y)$. For the class as a whole we shall have

$$S_1 = n^A{}_I\, s_A + n^B{}_I\, s_B \;; \quad S_2 = n^A{}_{II}\, s_A + n^B{}_{II}\, s_B \,. \tag{11}$$

If s_A and s_B are given, then, because of equation (8), S_1 and S_2 are known functions of time, $S_1(t)$, $S_2(t)$. We may now obtain x_1, x_2, y_1, and y_2 as functions of time, by using equations (4), in which u_1, u_2, v_1, and v_2 are derived from the corresponding $S_1(t)$ and $S_2(t)$, and in which the values a_1, a_2, b_1, and b_2, given by equations (9) and (10), are used. In this way we obtain $x_1 + y_1$ and $x_2 + y_2$ as functions of time. We thus may study the variation of the characteristic of the two classes with respect to time.

In a previous paper (2), we have derived a condition for one class to control the activities of the other. Denoting by N_1 the total population of class I, and by N_2 the total population of class II, that is, in our case

$$N_1 = n^A{}_I + n^B{}_I \;; \quad N_2 = n^A{}_{II} + n^B{}_{II} \,, \tag{12}$$

we found that under certain assumptions class I controls class II, if

$$N_1 > \frac{g_2}{g_1 + g_2}\,(N_1 + N_2). \tag{13}$$

Condition (13) is obtained from condition (8) of a previous paper (2) by changing the notations, namely by putting there $y_0 = 0$, $x_0 = N_1$, $a = g_2$, $a_0 = g_1$. The constants g_1 and g_2 depend, as we have seen before (loc. cit.), among other things on technical facilities possessed by the different classes. In general, therefore, g_1 will be a function of $(x_1 + y_1)$, while g_2 will be a function of $(x_2 + y_2)$. Hence

$$g_1 = g_1(t) \;; \quad g_2 = g_2(t). \tag{14}$$

In the inequality (13) all quantities involved are thus, for a prescribed shape of $s_A(x, y)$ and $s_B(x, y)$, as well as of $g_1(x_1 + y_1)$, $g_2(x_2 + y_2)$, known functions of time. We may investigate, then, for what times the relation (13) holds or ceases to hold, by substituting an equality sign for the inequality and by solving with respect to t.

Thus it may happen that inequality (13) holds for $t=0$ but ceases to hold for $t > t_1 > 0$. This would mean that class I will control class II from $t=0$ to $t=t_1$, but for $t > t_1$, the situation will change in accordance with equations developed before (2). Or inequality (13) may not hold for $t < t_1$ but hold for $t > t_1$. In that case class I becomes the controlling class at $t = t_1$. Or else the equation obtained from (13) may have several roots: t_1, t_2, etc., in which case we shall have a fluctuation of the control between the two classes. The explicit form of the equation for t_1 depends on the two mathematical assumptions: the shape of $s_A(x, y)$ and $s_B(x, y)$, and the shape of g_1 and g_2 as functions of $(x + y)$. The latter assumption is more of a physical nature. It depends largely on the nature of the quantities x and y — whether they are, for instance, material commodities, money, or some abstract knowledge. The simplest assumption is to make the g's proportional to $x + y$, but other assumptions are also possible. As to the choice of the functions $s_A(x, y)$ and $s_B(x, y)$, this is determined by purely psychophysical assumptions. We may have either the form suggested by Thurstone or some other plausible forms. Keeping a given functional expression for the g's and making different assumptions about the satisfaction function, we shall find different expressions for the variation of the interaction of social classes with respect to time. Thus we may study the effect of such a purely psychological factor as the satisfaction function upon the dynamics of society. For each choice of $s_A(x, y)$ or $s_B(x, y)$ and of the g's we have a definite mathematical problem, and the different problems thus obtained should be studied separately.

REFERENCES

1. Evans, Griffith C. Mathematical introduction to economics. New York: McGraw-Hill, 1930.
2. Rashevsky, N. Studies in mathematical theory of human relations. *Psychometrika,* 1939, 4, 221-239.
3. Rashevsky, N. Studies in mathematical theory of human relations, II. *Psychometrika,* 1939, 4 283-299.
4. Rashevsky, N. Studies in mathematical theory of human relations, III. Psychometrika, 1940, 4, 203-210.
5. Thurstone, L. L. The indifference function. *J. soc. Psychol,* 1931, 2, 139-167.

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1940

MAXIMUM LIKELIHOOD ESTIMATION AND FACTOR ANALYSIS

GALE YOUNG
OLIVET COLLEGE

Fisher's method of maximum likelihood is applied to the problem of estimation in factor analysis, as initiated by Lawley, and found to lead to a generalization of the Eckart matrix approximation problem. The solution of this in a special case is applied to show how test fallability enters into factor determination, it being noted that the method of communalities underestimates the number of factors.

I

Any observed score s may be regarded as the sum of a "true" score t and an "error" score x drawn from an error population with a distribution function $p(x)$. A common assumption in test theory is that the errors are normally distributed, i.e., that

$$p(x) = \frac{h}{\sqrt{\pi}} e^{-(hx)^2} , \tag{1}$$

where $h = 1/\sigma\sqrt{2}$, σ being the standard deviation.

Similarly, in factor analysis an observed score matrix $S = (s_{ij})$ may be regarded as arising from a true score matrix $T = (t_{ij})$ of lower rank r through the operation of independent random errors x_{ij}. There then arises the problem of estimating T from S, and Lawley (6) has recently suggested doing this by means of Fisher's method of maximum likelihood,* which is useful in the construction of estimates with desirable properties(2, 3, 4). The method consists in forming the probability density of occurrence of S from an arbitrary T, and then taking as an estimate that value of T which maximizes this density. With errors distributed as in (1), we are to maximize

$$L = \prod_{i,j} h_{ij} \, e^{-(h_{ij} x_{ij})^2} \tag{2}$$

subject to $T = S - X$ being of rank r. This is equivalent to maximiz-

* Dr. George Brown of Princeton University has independently made the same suggestion in some unpublished work.

ing log L, and the problem thus reduces to that of minimizing

$$\sum_{i,j} (h_{ij} x_{ij})^2 . \tag{3}$$

If the dispersions σ_{ij} of the error populations are equal, this is the same as the earlier "matrix approximation" procedure of Eckart (1); otherwise Eckart's problem is here generalized by the appearance of weighting factors h_{ij} for the residuals.

II

A solution of the general problem represented by (3) is not known to the present writer, but if the weighting matrix $H = (h_{ij})$ is of rank one the problem may be readily reduced to the unweighted Eckart problem, the solution of which is known.

If H is of rank one its elements are of the form $h_{ij} = a_i b_j$. If any element of H were zero, so then would be all others in either the same row or in the same column, and this row or column would drop out in (2) leaving a smaller matrix to work with. Hence if we assume H to be of rank one it is no further loss of generality to suppose its elements all non-zero.

Let the subscript 1 on a matrix denote the matrix obtained by weighting it with H, i.e., by multiplying each element by the corresponding element of H. Thus X_1 is the matrix whose elements are squared and added in (3). For H as specified above we have

$$\begin{aligned} X_1 &= A\,X\,B \\ S_1 &= A\,S\,B \\ T_1 &= A\,T\,B , \end{aligned} \tag{4}$$

where A and B are non-singular diagonal matrices with elements a_i and b_j. Therefore the weighting transformation does not in this case alter the rank of a matrix, and neither does the "unweighting" or inverse transformation given by $X = A^{-1} X_1 B^{-1}$, etc. Then *the solution of the weighted approximation problem* (3) *is obtained by weighting* S *to get* S_1, *finding the best rank* r *unweighted approximation* T_1 *to* S_1, *and then unweighting* T_1 *to get* T. For if T^* were a better rank r weighted approximation to S than is the matrix T thus obtained, then T^*_1 would be a better rank r unweighted approximation to S_1 than is T_1 because the quantity to be minimized is the same in either problem.

This proof does not hold for an arbitrary H, since weighting a matrix may then change its rank.

III

In practice one does not usually have data on the σ_{ij}, the dispersions of the various individuals j in the various tests i. If the test reliabilities r_i are available, a rough estimate of the σ_{ij} might be obtained by neglecting the differences from individual to individual and considering $\sigma_{ij} = \sigma_i$ to depend only upon the tests. Then, denoting the square of the length of the observed test vector by $l_i^2 = \sum_{j=1}^{N} s_{ij}^2$, one might take $N \sigma_i^2 = (1 - r_i)\, l_i^2$, or some similar expression, and thus obtain values for the σ_i. This leads to a rank one weighting matrix H, and thus comes under the case handled above. The same is true if σ_{ij} is the product of a quantity measuring the "unreliability" of the test and a quantity measuring the "unreliability" of the person: if values for σ_{ij} were actually known, one might take the best rank one approximation to the matrix (σ_{ij}) to get such a form.

The rough σ values obtained above give weighting factors

$$h_{ij} = h_i = \frac{1}{l_i\sqrt{1 - r_i}}, \tag{5}$$

and it is then seen that in weighting S to get the matrix S_1 upon which the actual work of unweighted approximation is performed we are in effect normalizing the test vectors to unit length, and then stretching them by a factor of $1/\sqrt{1 - r_i}$. Thus tests with lower reliabilities come out to have shorter vectors in the S_1 matrix.

This is reminiscent of the familiar "communalities" procedure, wherein the test vectors are normalized to unit length and then shortened by an amount corresponding to the error variance; in the present case the decrease would be to length $\sqrt{r_i}$. There is an important difference, however, in that the present process does not change the direction of the test vectors* and hence, over a considerable range of reliability, does not appreciably affect the dimensionality of the configuration, whereas the communality shortening is done without corresponding reduction in the off-diagonal scalar product elements of the correlation matrix and hence swings the test vectors all in toward each other, thus artificially increasing the clustering (7) and underestimating the number of factors. As an example, if two vectors making an angle of 60° are shortened to 0.707 of their original lengths, while their scalar product is kept unaltered, they will be

* In particular, if the tests have equal reliabilities, factoring the correlation matrix with unity in the diagonals is indicated.

swung completely into line. Similarly, a length reduction to 0.84 will align vectors originally 45° apart, and a reduction to 0.93 is sufficient to align vectors which were 30° apart.

IV

In this connection it may be noted that the practice of subtracting the error variance from the test variance is based essentially upon identifying a single observed value for the square of the length of the test vector with its expected value in an infinite sample. As a simple illustration, with errors distributed as in (1), the expected value of s^2 is given by

$$E(s^2) = t^2 + \sigma^2 . \tag{6}$$

If this be equated to the square of an observed value s_0 the result is

$$t^2 = s_0{}^2 - \sigma^2 , \tag{7}$$

which corresponds to the communality procedure.

On the other hand the method of maximum likelihood yields as the best estimate of t^2 simply $s_0{}^2$ itself.

V

(*Section added in response to comments by a referee*). A number of writers on the factor problem, e.g., Lawley, have formulated it as if different individuals taking a set of k tests were merely drawing samples from *the same* k-way distribution of variables in normal correlation. Such a distribution is specified by the means and variances of each test and the covariances of the tests in pairs; it has no parameters distinguishing different individuals. Such a formulation is therefore inappropriate for factor analysis, where factor loadings of the tests and of the individuals enter in symmetric fashion in a bi-linear form. It would perhaps be more suitable for psychophysics, where the differences between individuals are ignored and attention is focused on the test objects presented to them.

The point is that in factor analysis different individuals are regarded as drawing their scores from *different* k-way distributions, and in these distributions the mean for each test is the true score of the individual on that test. Nothing is implied about the distribution of observed scores over a population of individuals, and one makes assumptions only about the error distributions. In Section I above the errors were supposed to be normally distributed and uncorrelated; other assumptions would lead to expressions other than (2) for the likelihood function.

In section II the elements of the T matrix are obtained as certain functions of the elements of the S matrix, the nature of which may be concisely indicated by

$$T_1 = R\,S_1\,, \tag{8}$$

where (5) the matrix R involves the characteristic vectors of the matrix $S_1S'_1$. Under the present assumption that the elements of S and hence those of S_1 are distributed normally, it is seen that those of T_1 and T have a more complicated distribution because the elements of R themselves depend on S. To a first approximation, however, we may neglect this last dependence and thus have the t_{ij} as linear functions of the s_{ij}, with coefficients which become known when the calculations of section II have been carried out on an observed score matrix S. To this approximation the elements of T are distributed normally with variances expressed in terms of the σ_{ij} of the error distributions and the coefficients of the just-mentioned linear functions. One can then state the probable error of the t_{ij} estimates or proceed further to the construction of fiducial confidence belts.

The writer wishes to express his thanks to Dr. A. S. Householder for helpful discussion of this problem.

This work was aided in part by a grant from the Dr. Wallace C. and Clara A. Abbott Memorial Fund of the University of Chicago.

REFERENCES

1. Eckart, Carl and Young, Gale. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211-218.
2. Fisher, R. A. On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. A.*, 1921, 222, 309-368.
3. Fisher, R. A. Theory of statistical esitmation. *Proc. Camb. Phil. Soc.*, 1925, 22, 700-725.
4. Fisher, R. A. Inverse probability. *Proc. Camb. Phil. Soc.*, 1930, 26, 528-535.
5. Householder, A. S. and Young, Gale. Matrix approximation and latent roots. *Amer. Math. Monthly*, 1938, 45, 165-171.
6. Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. of Edinburgh*, 1939, 60, Part 1, No. 6.
7. Young, Gale. Factor analysis and the index of clustering. *Psychometrika*, 1939, 4, 201-208.

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

A MACHINE METHOD FOR COMPUTING THE BISERIAL CORRELATION COEFFICIENT IN ITEM VALIDATION

ELMER B. ROYER*

A method for computing the biserial correlation coefficients with the aid of punch card equipment is oultined. A numerical example and a work sheet layout is included in the presentation.

Pearson gives the following formula for biserial correlation,

$$r_{biserial} = \frac{M_1 - M_T}{\sigma_T} \cdot \frac{p}{z}, \tag{1}$$

where M_T = Mean of the total distribution of the Y-variable

M_1 = Mean in the Y-variable of category-one of the dichotomized or X-variable

σ_T = Standard deviation of the total distribution of the Y-variable

p = Proportion of cases included in category-one of the X-variable

z = The ordinate of the normal curve at the point of dichotomy.

The Pearson formula is well suited to the computation of biserial correlations in item validation, since the mean and standard deviation of the criterion variable need be computed only once for computing any number of item validity coefficients. In this case a slight change yields an algebraically equivalent formula suited to item validation by punched card treatment:

$$r_{biserial} = \frac{\Sigma Y_p - p\Sigma Y_t}{z} \cdot \frac{1}{N\sigma_t} \tag{2}$$

Where Σy_p = Sum of the criterion scores of those passing the item

* This method was developed by Dr. Elmer B. Royer prior to his untimely death on April 3, 1939. It has been prepared for publication by some of his former associates, not only as a contribution to science but also as a tribute to the memory of a true scientist.

$$p = \text{Proportion of cases passing the item} = \frac{N_1}{N}$$

$$\Sigma Y_t = \text{Sum of criterion scores of the total group}$$

$$N = \text{Number of cases in the total group}$$

(σ_T and z have the same meanings as above, and N_1 is the number of persons responding successfully to the item.)

Items are scored: Fail = 0; Pass = 1.

In the above expression, the term $\dfrac{1}{N\sigma_T}$ obviously is a constant in the computation of the validity coefficients for a given population and criterion.

Coding the Data

Assuming that 150 two-response items have been given to 298 subjects, the correct responses may be punched in 50 columns of an 80-column card. The coding may be set up somewhat in this fashion:

Cols. 1 to 10 — Identifying material.

" 11 and 12 — Criterion scores, (Y).

" 13 to 62 — Item responses. Punches in row I indicate successful responses only for items 1 to 50; in row 3, for items 51 to 100; in row 5 for items 101 to 150. Thus, if 1 is punched in column 14, it indicates that item 2 has been correctly answered; and if 3 is not punched in column 15, it indicates that item 53 has been answered incorrectly, etc. The coding of item responses may vary with the number of items and responses. The only requisite is that it be easily possible to sort out the group making a particular response and subsequently to tabulate the criterion scores of that group.

Cols. 63, 64, 65, and 66 — The reciprocal of N, in this case .003356, which is punched without the initial zeros.

Col. 67 — Punch "1" in each card for use in obtaining an item count.

TABULATION — Steps in finding $\sum Y_t$ and σ_t

Assuming that there are scores of large magnitude in the criterion variable, the digiting method of obtaining the summations (implying that $Y = 10a + b$, where a = tens-digit and b = units-digit) is employed.

1. Sort on units column of criterion scores, and arrange cards in descending order of value.
2. Set tabulating machine to control on units column and to add and cumulate criterion scores. Print Y-units column in the list bank and print the item count and criterion scores as progressive totals. (See tabulation below.)
3. Sum the cumulated Y-scores column, omitting the zero-class entries. Enter this number, 56996, on the work sheet where it reads "$\sum$ units."
4. The highest value in the cumulated scores column will be $\sum Y_t$. Enter this number on the work sheet in the place designated.
5. Repeat steps 1, 2, and 3, sorting instead on the tens-column of criterion scores.
6. Enter "$\sum$ tens" on work sheet. Check $\sum Y_t$.
7. Enter N on work sheet.

The tabulation sheets will appear as follows :

Units Column of Y	f (Item-Count) Cumulated Frequency		Cumulated Y Scores	
9	25		1165	
8	53		2799	
7	86		4230	
6	119		5708	
5	147		6448	
4	171		7494	
3	209		8798	
2	239		9768	
1	267	$1316 = \Sigma f$ (units)	10586	56996 = "Σ units"
0	$298 = N$		$11446 = \Sigma Y_t$	

Tens Column of Y	f (Item Count)		Cumulated Y Scores	
9	46		4381	
8	77		7717	
(7)*	(77)*		(7717)*	
6	125		9091	
5	109		9256	
4	110		9298	
3	116		9480	
2	143		10116	
1	210	$1013 = \Sigma f$ (tens)	11034	$78090 =$ "Σ tens"
0	$\overline{298} = N$		$\overline{11446} = \Sigma Y_t$**	

*—Since there are no 7's, the respective extensions of the 8-row must be added twice as indicated by the parentheses.

**—Check: $\Sigma Y_t = 10 \Sigma f$ (tens) $+ \Sigma f$ (units)
$11446 = 10\ (1013) + 1316$

All the values needed to compute the constant $1/N\sigma_T$ are now known. The operations yielding this value are indicated on the work sheet.

To obtain r_{bis} for any item, sort out the cards of those persons passing the item. Place these in the hopper of the tabulator and set the machine to add on the criterion scores, the reciprocal of N, and the item count. The tabulation of five items would appear:

Item	Item Count	ΣY_p	p
1	148	4401	496688
2	106	5130	355736
3	223	9421	748388
4	78	2900	261768
5	26	1160	087256

Enter $\sum Y_p$ and p on the work sheet, and find z in the Kelly-Wood tables and record. Follow the operations indicated by the work sheet. The computation of r_{bis} for the five items above is shown on the work sheet.

WORK SHEET
Computation of r_{BIS}

$$r_{BIS} = \frac{\Sigma Yp - p\Sigma Y_T}{z} \cdot \frac{1}{N\sigma_T}$$

Computation of Constant

Operation		Result
Σf	298	N
Σ units	56,996	
Σ tens	780,900	
Σ hundreds	00	
Add	837,896	ΣY^2_T
$N.\Sigma Y^2_T$	249,693,008	$N\Sigma Y^2_T$
ΣY_T	11446	ΣY_T
$\Sigma Y_T \cdot \Sigma Y_T$	131,010,916	$(\Sigma Y_T)^2$
$N\Sigma Y_T^2 - (\Sigma Y_T)^2$	118,682,092	$(N\sigma_T)^2$
$\sqrt{(N\sigma)^2}$	10894.1	$N\sigma_T$
$1/N\sigma_T$	.000091793	Constant Multiplier
$\frac{\Sigma Y_T}{N}$	38.41	M_T
$\frac{N\sigma_T}{N}$	36.56	σ_T

*Check on item 1 computations
$\Sigma Yp - p\Sigma Y_T = (1-p)\Sigma Y_T - (\Sigma Y_T - \Sigma Y_p)$
$-1284.228 = (1 - .4967)11446 - (11446 - 4401)$
$-1284.228 = .5033 \times 11446 - .7045$
$-1284.228 = -1284.228$

Item	ΣYp	p	Z	$A = \Sigma Yp - p\Sigma Y_T$	$B = A/Z$	$r_{BIS} = B \cdot 1/N\sigma_T$
1	4401	.4967	.3989	–1284.228*	–3219.423	–.296
2	5130	.3557	.3726	1058.658	2841.272	.261
3	9421	.7484	.3189	854.814	2680.508	.246
4	2900	.2618	.3255	– 96.563	– 296.661	–.027
5	1160	.0873	.1587	160.764	1013.006	.093
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						

REFERENCES

1. Dunlap, J. W. Computation of descriptive statistics. New York: Ralph C. Coxhead Corporation, 1937, 78-81.
2. Dunlap, J. W. and McNamara, W. J. A graphical method for computing the standard error of biserial *r*. *J. exper. Educ.*, 1934, 2, 274-277.
3. Pearson, K. On a new method of determining correlation between a measured character *A* and a character *B*, of which only the percentage of cases wherein *B* exceeds of falls short of a given intensity is recorded for each grade of *A*. *Biometrika*, 1909, 7, 96-105.

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

1. [illegible]
2. [illegible]
3. [illegible]

PSYCHOMETRIKA—VOL. 6, NO. 1
FEBRUARY, 1941

MUSIC ABILITY

J. E. KARLIN
UNIVERSITY OF CHICAGO

Two batteries of music tests were factored by the centroid method. From each battery three oblique factors were extracted and in each case were tentatively identified as tonal sensitivity, retentivity (memory for elements), and memory for form. The correlations of the music tests of one battery with subtests of Cattell's intelligence test and with tests of a literary nature are also reported.

It has been held that musical ability is some dominantly unitary feature of mental life incapable of analysis into simpler components by rational methods. This study is a preliminary investigation of the music field by Thurstone's method of multiple factor analysis.

An account is given of an analysis of two batteries of music tests typical of the music test literature. The first battery was assembled by the present writer,* the second by Drake.† The battery given by the writer to 120 undergraduate students in the University of Cape Town (South Africa) consisted of 19 tests, as follows:

Tests 1 to 6 were the six parts of R. B. Cattell's intelligence test, scale III; tests 7 to 9 inclusive were of a literary nature; tests 10 to 19 were chosen as music tests. The origin of each music test is indicated in Table 1, those without any specific indication having been devised either by the writer or by members of the faculty of music in the University of Cape Town after the pattern of the Seashore tests which were themselves deemed too difficult for the population to be tested. It was hoped that the inclusion of intelligence tests in the same battery as the music tests would throw light on the much debated question of cognition in music. Similarly, the literary tests could perhaps provide the first steps towards possible evidence of a general artistic factor. The table of intercorrelations, computed as Pearson product-moment coefficients, is reproduced in Table 1 and is immediately seen to be highly informative on both the foregoing points. While the literary and intelligence tests correlate highly with each other and among themselves, of the 90 intercorrelations between

* Karlin, J. E. A multiple factor analysis of musicality. M. A. thesis, University of Cape Town, 1939.

† Drake, R. M. A factorial analysis of music tests by the Spearman tetrad-difference technique. *J. Musicology*, 1939, 1, 1.

the music tests and the rest of the battery, only 25 were as high as .10 and the mean correlation was only .05.

TABLE I

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	.																		
2	.512																		
3	.409	.319																	
4	.232	.334	.450																
5	.346	.269	.411	.413															
6	.299	.356	.391	.493	.333														
7	.391	.264	.285	.336	.437	.265													
8	.564	.270	.493	.476	.502	.451	.476												
9	.396	.353	.400	.429	.430	.354	.432	.562											
10	−.056	−.069	.025	−.052	−.096	−.099	.171	−.005	−.084										
11	.072	.096	.013	−.017	.022	−.030	.059	.026	−.072	.060									
12	−.114	−.055	−.049	−.059	−.025	−.073	.020	−.025	.081	.343	.122								
13	−.110	.028	−.130	−.061	−.020	−.034	−.034	−.002	−.075	.061	.001	.095							
14	.138	.204	.098	.163	.094	.206	.044	.281	.095	.110	.072	.210	.110						
15	−.091	−.042	.069	.130	−.124	−.014	.034	−.064	.004	.359	.203	.439	.062	.228					
16	.088	.057	.067	.096	.023	−.036	−.060	.137	.049	.269	.250	.267	.191	.331	.592				
17	.161	.059	.016	.018	.124	.131	.141	−.084	.029	.227	.262	.389	−.011	.322	.342	.510			
18	.062	.235	−.079	.007	−.079	−.082	−.120	−.002	−.105	−.009	.093	.067	.059	.002	.056	−.025	−.056		
19	.066	.016	.127	.169	.152	.080	.168	.324	.285	.176	.032	.091	−.005	.218	.180	.326	.231	−.023	

1. Synonyms
2. Classification
3. Opposites
4. Analogies
5. Completion of sentences
6. Inferences
7. Reading comprehension
8. Vocabulary
9. Poetical appreciation
10. Pitch discrimination
11. Tonal memory
12. Interval discrimination
13. Consonance
14. Rhythm
15. Time
16. Musical Memory (Drake)
17. Retentivity (Drake)
18. Intensity (Seashore)
19. Emotional sensitivity

It might appear, then, that musical ability pertains largely to a field of its own. Yet it may be unwise to conclude this too hastily. With the isolation of the primary mental abilities by Thurstone, the concept of general intelligence or g, as Spearman put it, is becoming less and less widely accepted as meaningful in any unitary sense. It may be, however, that there is over-lap between the more elemental components of intelligence and fundamental abilities peculiar to the music domain. Likewise the lack of correlation between the music tests and the literary tests indicates the closeness of identity of the verbal factor with some aspect of general intelligence.

The battery was accordingly split into the musical and non-musical halves, and the music battery of ten tests, that is, tests 10 to 19 inclusive, was factored by the centroid method. After three factors were extracted, the median residual coefficient was .034 and the median probable error corrected for attenuation .035, and the highest residual was .091. The residue was therefore deemed unsystematic. Table 2 shows the rotation of the three centroid vectors to the three primary vectors of the present system in accordance with the demands of simple structure. Table 3 gives the correlations of the pri-

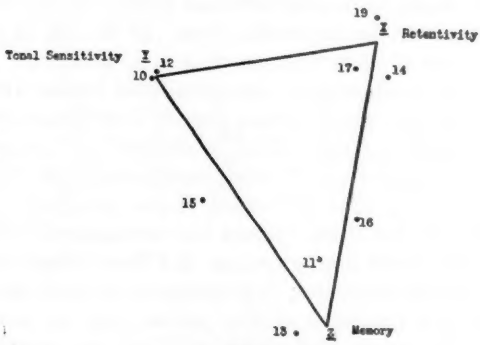

FIGURE 1

maries. Figure 1 is a pictorial representation of the trait configuration in relation to the primary vectors as derived by the method of extended vectors*.

TABLE 2

	F_c			Λ			F_r		
	I	II	III	X	Y	Z	X	Y	Z
10.	.448	—.281	.148				—.002	.431	.005
11.	.311	.073	—.169				.048	.013	.290
12.	.564	—.347	.190				.004	.537	.005
13.	.173	.021	—.156				—.024	.017	.214
14.	.443	.241	.150	.287	.333	.406	.398	—.057	.082
15.	.697	—.246	—.187	.768	—.933	.153	—.096	.437	.414
16.	.732	.291	—.243	.573	.133	—.901	.294	—.060	.561
17.	.625	.237	.236				.497	.018	.077
18.	.059	—.110	—.177				.017	.099	.167
19.	.314	.150	.197				.318	—.009	—.027

TABLE 3

$\Lambda' \Lambda$

	X	Y	Z
X	1.000		
Y	—.545	1.000	
Z	—.282	.112	1.000

Two observations are called for:

1. Test 18 (Intensity) is omitted from consideration since it showed negligible correlation with the rest of the battery almost throughout, having a communality of .047.
2. The condition of a positive manifold is fulfilled.

* Thurstone, L. L. A new rotational method in factor analysis. *Psychometrika*, 1938, 3, 199-218.

The only claim made regarding musical ability in this battery is that the variance present in the tests, from .10 to .68, is plausibly explained here in terms of three factors. It will be necessary to devise tests which will have much higher communalities before it becomes evident how many psychological factors are involved in any such battery. It is unlikely that musical ability in general can be reduced to only three functional unities. With such small batteries, the insufficiency of data allows only of a somewhat vague structure; although the general character of the three factors can be seen with reasonable assurance in that the trait configuration did functionally outline a three-dimensional simple structure, it is necessary to have many more tests in order that the positions of the planes may be determined. With further study, the number of planes will be more exactly defined so as to give a multi-dimensional system, which may or may not be orthogonal. The present system is oblique, but it may well be that as the parameters become overdetermined by test data the planes will define themselves as being orthogonal. In the present case, factor Y appears to be some sort of *tonal sensitivity* factor, having its greatest weight on tests 10 and 12; factor X seems to be a *retentivity* or memory for elements factor with highest load on test 17; factor Z is a memory for form factor with maximal saturation in test 16. The two memory factors are obscure in outline apart from their retentive nature. It is imperative that future work be directed towards devising many further tests which will serve to accentuate the planes in general and the corners of the structure in particular.

A very similar procedure was adopted for a reanalysis of music test data assembled by Drake. His table of raw coefficients is reproduced in Table 4.

TABLE 4

		1	2	3	4	5	6	7	8
Musical memory	1.								
Pitch	2.	.466							
Retentivity	3.	.456	.311						
Rhythm	4.	.441	.296	.185					
Intensity	5.	.375	.521	.184	.176				
Time	6.	.312	.286	.300	.244	.389			
Tonal movement	7.	.247	.483	.378	.121	.211	.210		
Tonal memory	8.	.207	.314	.378	.341	.153	.289	.504	

Again three factors were all that could be extracted, with the median residual coefficient .036 and the median probable error corrected for attenuation .029. With only 8 tests, a fourth factor is not justified. The rotation of the axes is shown in Table 5 and the correlation of the primaries is given by Table 6.

TABLE 5

	F_c			Λ			F_r		
	I	II	III				X	Y	Z
1	.643	—.258	—.202				.510	.073	—.039
2	.692	—.138	.346	X	Y	Z	.043	.572	.136
3	.573	.220	—.198	.480	.323	.340	.427	—.043	.384
4	.486	—.169	—.323	—.096	—.203	.935	.531	—.108	—.028
5	.547	—.369	.283	—.873	.926	.085	.051	.513	—.134
6	.523	—.100	—.075				.326	.120	.078
7	.575	.386	.274				.000	.361	.580
8	.582	.378	—.109				.338	.010	.542

TABLE 6

$\Lambda' \Lambda$

	X	Y	Z
X	1.000		
Y	—.634	1.000	
Z	—.001	—.154	1.000

From Figure 2 it would appear that here too a three-dimensional oblique simple structure prevails. Factor Y looks very much like the *tonal sensitivity* factor already identified with highest load on test 2; factor X is a *memory* factor with heavy loads on tests 1, 3, and 4; factor Z is probably the *retentivity* factor, it being most evident in tests 7, 8, and 3.

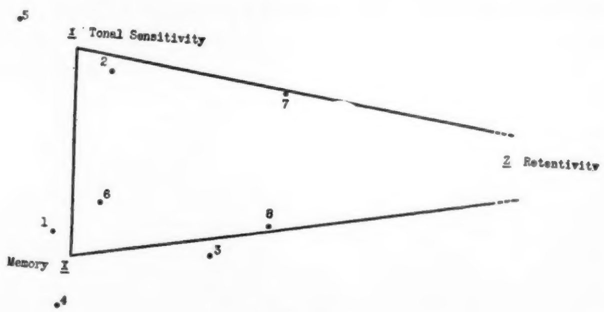

FIGURE 2

The same warnings must be given here as were appropriate in the previous analysis. The agreement between the results of the two analyses is promising for further and more extensive studies. Such studies are in progress at the present time. Their ultimate purpose is the isolation of the primary musical abilities.